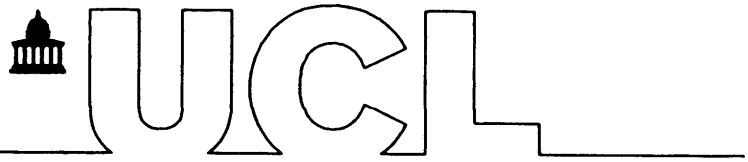
UNIVERSITY OF LONDON THESIS

Degree PhD Year 2007 Name of Author TINA SHAH



Linking observational and genetic approaches to determine the role of C-reactive protein in coronary disease risk.

Tina Shah

Departments of Clinical Pharmacology & Cardiovascular Genetics, University College London

Thesis submitted for the degree of PhD, 2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

Acknowledgements

I would like to acknowledge the help and support of everyone who has made this thesis possible, especially my supervisors, Dr Aroon Hingorani and Professor Steve Humphries, who have provided constant support and guidance. I would also like to thank the British Heart Foundation, who generously funded this project.

I also wish to thank all those with whom I have collaborated in this thesis, especially Dr. Juan Pablo Casas, Ms. Jackie Cooper, Dr. Francesco D'Aiuto and Dr. Leonelo Bautista. I would also like to acknowledge those who have provided me with the data sets used in this thesis, in particular Dr Manjinder Sandhu, Professor Kay-Tee Khaw and Professor Nick Wareham, as well as those who have provided help with the methodology used in this thesis, especially Ms. Jutta Palmen, Ms. Kerra Pearce and Dr. Lorna Tinworth.

Finally, I would like to thank my family, who have provided unfailing support throughout this time, and I would like to dedicate this thesis to them.

Abstract

Abstract

Inflammation may contribute to atherogenesis. C-reactive protein (CRP) is an acute phase protein whose circulating concentrations provide an index of infection or inflammation. Prospective studies have consistently shown associations of CRP among apparently healthy individuals with risk of future hypertension, diabetes and coronary heart disease (CHD) events, leading to the proposal that CRP may play a causal role in atherogenesis, and its measurement could help predict CHD events.

However, CRP is associated with a range of risk factors and biomarkers linked to atherosclerosis, so its predictive utility could be limited once these are accounted for. Its causal relevance is also uncertain because of the potential for confounding and reverse causation in observational studies. Experimental data have not resolved controversy on causation, as recent studies have indicated that apparently pro-atherogenic actions of CRP may have been mediated by preservatives or contaminants in commercial CRP preparations. Studies were therefore undertaken to evaluate the role of CRP in the prediction and pathogenesis of atherosclerotic events, by linking observational with genetic approaches.

Appropriate analytical tools were used to evaluate the performance of CRP as a screening test for CHD. Single nucleotide polymorphisms (SNPs) in the CRP gene were identified using a bioinformatic approach and linkage disequilibrium was assessed to identify "tagging SNPs" from which haplotypes could be inferred. A literature-based systematic review was undertaken to obtain precise estimates of the effect of SNPs on CRP concentration and to confirm the expected balanced distribution of potential confounders. In new studies, associations between haplotype (or genotype for a potentially functional promoter variant), and CRP concentration were tested. The effect of SNPs and haplotypes on CRP following an acute inflammatory stimulus was also assessed. The frequency of CRP SNPs and their effect on CRP were also examined in different ethnic groups. Having confirmed that SNPs and haplotypes of the CRP gene provide an unbiased proxy for CRP itself, associations with pro-atherogenic phenotypes, incident hypertension, diabetes and CHD were tested.

Despite showing association with CHD, CRP performed poorly as a predictive test. Reasons for the poor performance were explored and comparisons were made with established risk factors and emerging biomarkers. SNPs and haplotypes were associated robustly with CRP, but not with other risk factors for CHD. A Mendelian randomisation approach linking genetic and non-genetic information provided no positive evidence that CRP induces a proatherogenic phenotype, nor increases risk of hypertension, diabetes or CHD. However, larger sample sizes will be required to exclude a small but potentially important causal influence of CRP on

cardiovascular risk. In the absence of a selective CRP-lowering drug, this approach provides the only currently practical method for obtaining randomised evidence on the causal relevance of CRP in atherogenesis.

Publications

Publications

D'Aiuto, F., Casas, J.P., Shah, T., Humphries, S.E., Hingorani, A.D., & Tonetti, M. 2005. C-reactive protein (+1444C>T) polymorphism influences CRP response following a moderate inflammatory stimulus. *Atherosclerosis* 179(2):413-417.

Casas, J.P., Shah, T., Cooper, J., Hawe, E., Stephens, J., Yudkin, J.S., Livingstone, S., Colhoun, H.M., Bautista, L.E., Meade, T., Gaffney, D., McMahon, A.D., Hamsten, A., Sattar, N., Humphries, S.E., & Hingorani, A.D. 2006. Insight into the nature of the CRP-coronary event association using Mendelian randomization. *International Journal of Epidemiology* 35(4):922-931.

Hingorani A.D., Shah T., Casas J.P. 2006. Linking observational and genetic approaches to determine the role of C-reactive protein in heart disease risk. *European Heart Journal* 27(11):1261-1263.

Shah, T., Cooper, J., Casas, J.P., Cavalleri, G.L., Pearce, K., Tinworth, L.E., Miller, G.J., Hingorani, A.D., & Humphries, S.E. C-reactive protein and coronary events. Re-evaluation of predictive utility and causal relevance. *(Submitted)*.

Shah, T., Casas, J.P., Cooper, J., Yudkin, J.S., Colhoun, H.M., Bautista, L.E., Hamsten, A., Sattar, N., Humphries, S.E., & Hingorani, A.D. CRP and CVD – Test for causality using Mendelian randomisation. *European Atherosclerosis Society 75th Congress 2005 (Poster presentation for Young Investigator Award)*.

Shah, T., Cooper, J., Casas, J.P., Cavalleri, G.L., Pearce, K., Tinworth, L.E., Miller, G.J., Hingorani, A.D., & Humphries, S.E. C-reactive protein in the prediction and pathogenesis of cardiovascular events. *British Atherosclerosis Society Spring Meeting 2006 (Oral presentation for Young Investigator Award finalist)*.

Declaration of work presented

Declaration of work presented in this thesis

I performed searches on public domain resources to generate a map of all common polymorphisms in the CRP gene, and I evaluated linkage disequilibrium between variants. I conducted a systematic review and meta-analysis of all published studies investigating associations of CRP genotype and CRP concentration. Where necessary, I contacted principal authors of these studies to obtain additional unpublished data. I conducted these meta-analyses myself. I also performed meta-analyses of published data investigating differences in CRP concentration in different ancestral populations. I designed primers for, and genotyped CRP polymorphisms in the NPHSII, Periodontal disease, ETNIAS, and LEADER studies. I also helped to genotype polymorphisms in the Ely study. All other genotyping was conducted in the respective centres for each study and the data sets were available for analysis. Using haplotype tagging software, I developed panels of tagging SNPs for CRP suitable for use in populations of European and non-European ancestry. For the Periodontal disease study, I carried out enzyme-linked immunosorbent assays (ELISAs) to measure IL-6 concentrations. I prepared tabular data for statistical analyses conducted by Ms. Jackie Cooper, Dr. Francesco D'Aiuto, Dr. Juan Pablo Casas, and Dr. Leonelo Bautista. I helped to plan these analyses and discussed their interpretation extensively with these collaborators. I used recombinant techniques to produce CRP gene promoter constructs in the pGL3 Basic vector to investigate whether CRP promoter variants might influence CRP mRNA transcription.

Abbreviations

Abbreviations

ACE	angiotensin converting enzyme
ACS	acute coronary syndrome
AHA/CDC	American heart association/ Centers for disease control and prevention
ANOVA	one-way analysis of variance
Apo	apolipoprotein
APRE	acute phase responsive element
ATP	adenosine triphosphate
AUC	area under receiver operating characteristic curve
BMI	body mass index
BP	blood pressure
BSA	bovine serum albumine
C. pneumoniae	*Chlamydia pneumoniae*
C/EBP	CCAAT-enhancer binding protein
CAD	coronary artery disease
CbVD	cerebrovascular disease
CEPH	Centre d'Etude du Polymorphisme Humain
CETP	cholesteryl ester transfer protein
CHD	coronary heart disease
CMV	cytomegalovirus
CREBH	cyclic AMP response element–binding protein H
CRP	C-reactive protein
CVD	cardiovascular disease
DIP	deletion/insertion polymorphism
DM	diabetes mellitus
dNTP	deoxyribonucleotide triphosphate
dUTP	deoxyuridine triphosphate
EBCT	electron beam computerised tomography
ECG	electrocardiogram
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EMSA	electrophoretic mobility shift assay
eNOS	endothelial nitric oxide synthase

FH	familial hypercolesterolaemia
H. pylori	*Helicobacter pylori*
H-APF	hepatocyte-acute phase factor
HCl	hydrogen chloride
HDL	high density lipoprotein
HMG-CoA reductase	3-hydroxy-3-methyl-glutaryl-CoA reductase
HNF	hepatocyte nuclear factor
HPLC	high performance liquid chromatography
HR	hazard ratio
HRT	hormone replacement therapy
ICAM-1	intercellular cell adhesion molecule-1
IHD	ischaemic heart disease
IL-1Ra	interleukin-1 receptor antagonist
IL-1β	interleukin-1β
IL-6	interleukin-6
IL-6RE	IL-6 responsive element
IMT	intima media thickness
KCl	potassium chloride
LD	linkage disequilibrium
LDL	low density lipoprotein
LDLR	low density lipoprotein receptor
LEAD	lower extremity arterial disease
LPS	lipopolysaccharide
LR	likelihood ratio
MADGE	microtitre array diagonal gel electrophoresis
MAF	minor allele frequency
MCP-1	monocyte chemoattractant protein-1
MDV	Marek's disease virus
$MgCl_2$	magnesium chloride
MHV-68	murine herpes virus 68
MI	myocardial infarction
MMP	matrix metalloproteinase
mRNA	messenger RNA
MTHFR	methyltetrahydrofolate reductase
NaCl	sodium chloride

NFIL-6α	nuclear factor interleukin-6α
NFκB	nuclear factor kappa-B
NIDDM	non insulin dependent diabetes mellitus
NO	nitric oxide
NOS	nitric oxide synthase
OR	odds ratio
ORF	open reading frame
P. gingivalis	*Porphyromonas gingivalis*
PAI-1	plasminogen activator inhibitor-1
PCh	phosphocholine
PCR	polymerase chain reaction
PCSK9	proprotein convertase subtilisin/kexin type 9
PGA	Programs for genomic applications
PPARγ	peroxisome proliferator-activated receptor-γ
PVD	peripheral vascular disease
RA	rheumatoid arthritis
RCT	randomised controlled trial
RFLP	restriction fragment length polymorphism
ROC curve	receiver operating characteristic curve
ROS	reactive oxygen species
RR	relative risk
SAA	serum amyloid A
SAP	serum amyloid P component
SD	standard deviation
SDS	sodium dodecyl sulfate
SE	standard error
SES	socioeconomic status
SLE	systemic lupus erythematosus
SMC	smooth muscle cell
SNP	single nucleotide polymorphism
SSCP	single strand confirmation polymorphism
STAT	signal transducers and activators of transcription
TBE	tris-borate buffer
TEMED	N,N,N'N'-Tetramethylethylene- diamine
TIMI	thrombolysis in myocardial infarction
TNF-α	tumour necrosis factor-α

tPA	tissue plasminogen activator
Tris	tris hydroxymethylaminoethane
Triton-X-100	octylphenol ethoxylate
USF-1	upstream transcription factor-1
UTR	untranslated region
VCAM-1	vascular cell adhesion molecule-1
VLDL	very low density lipoprotein
WHO	world health organisation

Contents

Contents

Introduction

1.1 Atherosclerosis

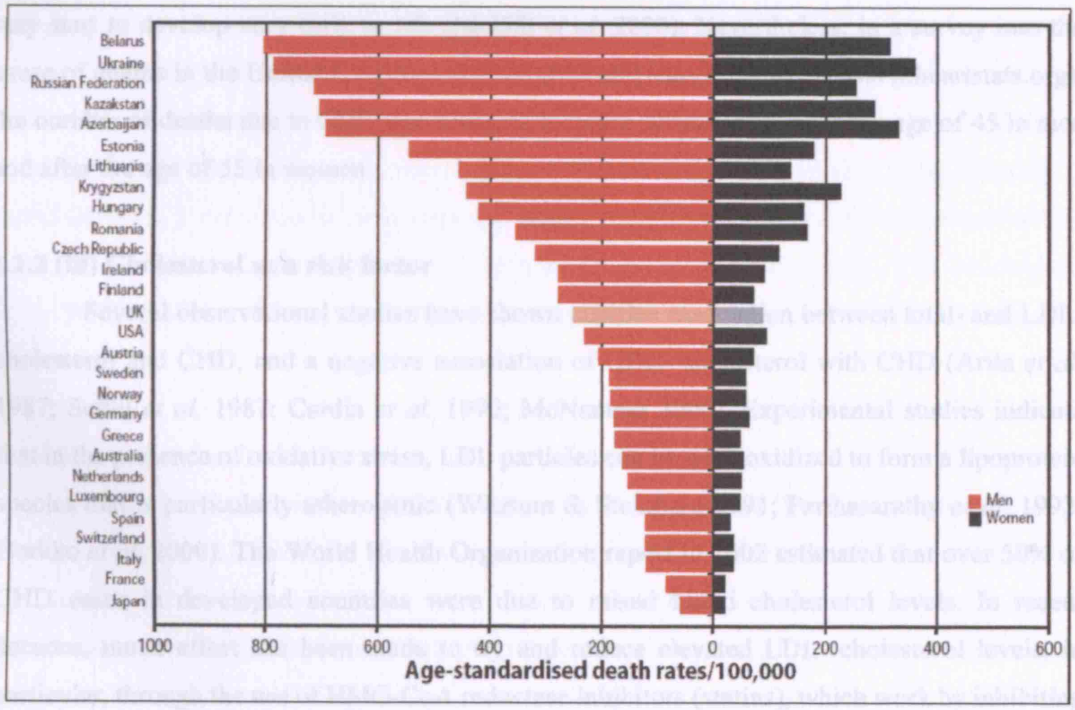
1.1.1 What is atherosclerosis?

Atherosclerosis is the process by which lipid-rich plaques accumulate in the intimal lining of the arterial wall, leading to a partial or complete obstruction of the blood flow through arteries in the heart, brain and lower limb vasculature. Atherosclerotic plaques are thus the substrate for cardiovascular disease (CVD), which includes coronary heart disease (CHD), a clinical end-point studied extensively in this thesis, as well as peripheral vascular disease (PVD) and stroke.

1.1.2 Epidemiology of cardiovascular disease

CVD is one of the main causes of premature death in the developed world, and its rates are increasing in developing countries (British Heart Foundation, www.heartstats.org). Coronary disease is the main cause of mortality in the UK, accounting for 39% of all deaths (see Figure 1.1).

Figure 1.1. Death rates from CHD in men and women aged 35-74, 1999 (data from the British Heart Foundation, www.heartstats.org).



A number of 'classical' or 'established' risk factors for CHD have already been established from population studies, molecular studies and randomised controlled trials (RCTs), which are considered to play an aetiological role in disease development. These risk factors include male gender (Wingard *et al.* 1983), age (Stary 1989), cholesterol (Wyndham *et al.* 1987; Gordon *et al.* 1981), blood pressure (Stamler *et al.* 1993), diabetes (Reaven 1988), smoking (Kannel *et al.* 1987), socio-economic status (Brunner *et al.* 1997), and birthweight (Barker 2004).

1.1.2 (i) Gender as a risk factor

It is well established that the prevalence of CHD is greater in men than in women until menopause (Lerner & Kannel 1986; Wingard *et al.* 1983). One explanation for this may be male-female differences in sex hormone profiles. A review of clinical studies in this area (Kalin & Zumoff 1990) has shown that there is a low prevalence of coronary disease in premenopausal women, with a loss of protection after menopause. Despite this, RCTs of hormone replacement therapy in postmenopausal women have not indicated protection from CHD, leading to ongoing uncertainty about the precise reason for male-female difference in CHD rates (Samsioe 2003).

1.1.2 (ii) Age as a risk factor

The risk of clinical coronary events such as angina, acute coronary syndrome and sudden death increases with age, but evidence from post-mortem studies of young accident victims, as well as recent intra-vascular ultrasound studies suggests that atherosclerosis itself may start to develop very early in life (McGill *et al.* 2000). Nevertheless, in a survey into the cause of deaths in the United Kingdom in 2002 (British Heart Foundation, www.heartstats.org), the number of deaths due to CHD was found to increase substantially after the age of 45 in men and after the age of 55 in women.

1.1.2 (iii) Cholesterol as a risk factor

Several observational studies have shown positive association between total- and LDL-cholesterol and CHD, and a negative association of HDL- cholesterol with CHD (Arita *et al.* 1987; Seftel *et al.* 1987; Cardia *et al.* 1990; McNamara 2000). Experimental studies indicate that in the presence of oxidative stress, LDL particles can become oxidized to form a lipoprotein species that is particularly atherogenic (Witztum & Steinberg 1991; Parthasarathy *et al.* 1992; Horkko *et al.* 2000). The World Health Organisation report in 2002 estimated that over 50% of CHD cases in developed countries were due to raised blood cholesterol levels. In recent decades, much effort has been made to try and reduce elevated LDL- cholesterol levels, in particular, through the use of HMG-CoA reductase inhibitors (statins), which work by inhibiting the rate-limiting enzyme in cholesterol biosynthesis, leading to a consequent up-regulation of surface LDL-receptors in hepatocytes that clear LDL- cholesterol particles from the circulation.

Multiple RCTs have been conducted (summarised in a recent meta-analysis; the Cholesterol Lowering Trialists Collaboration), which individually and in conjunction, demonstrate that statin therapy to reduce LDL- cholesterol is associated with a decrease in CHD events (Ross *et al.* 1999; LaRosa *et al.* 1999; Vrecer *et al.* 2003). Moreover, the benefits observed from cholesterol lowering in RCTs appear close to those expected from the prior observational studies, providing strong evidence that LDL- cholesterol is causal in atherosclerosis. Several secondary prevention trials, such as the Scandinavian Simvastatin Survival Study, and the PROSPER study have also been carried out to examine the effects of statins in patients with established CHD and have found a similar relative reduction in coronary events among the statin treated groups (Pedersen *et al.* 1998; Shepherd *et al.* 2002). The relationship between cholesterol and stroke is less certain, and may be complicated by a positive association with ischaemic stroke and a negative association with haemorrhagic stroke (Ebrahim *et al.* 2006).

1.1.2 (iv) Blood pressure as a risk factor

Another important risk factor for CHD is raised systolic and diastolic blood pressure (hypertension), which is associated both with CHD and stroke, with a graded and progressive relationship between blood pressure level and disease risk. The pathogenesis of hypertension is itself a multifactorial process that is likely to involve the interaction of genetic and environmental factors. In varying degrees, abnormalities of volume regulation, enhanced vasoconstriction, and remodelling of the arterial wall (decreasing lumen diameter and increasing resistance) contribute to the development of hypertension (Berk & Alexander 1995). In meta-analyses of observational studies, there is a continuous, graded, independent association between baseline levels of blood pressure in apparently healthy individuals and risk of later vascular events over the whole range of blood pressure values (Stamler *et al.* 1989; Lewington *et al.* 2002). RCTs of drugs to lower blood pressure have confirmed that such lowering is associated with a reduction in cardiovascular events with the consistency between observed and expected reductions suggesting a causal role in atherosclerosis (Collins *et al.* 1990; MacMahon *et al.* 2000; The Heart Outcomes Prevention Evaluation Study Investigators 2000; Lewington *et al.* 2002).

1.1.2 (v) Diabetes as a risk factor

Type II diabetes mellitus (T2DM) is another key risk factor for CHD. T2DM and lesser degrees of glucose intolerance commonly occur together with a collection of clinical and biochemical features, which have been called the metabolic syndrome (Reaven 1988; Wajchenberg *et al.* 1994). These features include central obesity, hypertriglyceridaemia, low serum concentrations of HDL- cholesterol, and high blood pressure. There is considerable evidence that resistance to insulin-stimulated glucose uptake leads to a compensatory increase in plasma insulin concentration, enhanced hepatic VLDL secretion, and hypertriglyceridemia,

although it is unclear if the association of triglycerides and coronary disease is causal (Pickup & Crook 1998). Association studies have also established a relationship between plasma insulin concentration and blood pressure. Epidemiological studies have found that patients with T2DM have a greater individual risk of developing atherosclerotic vascular disease (Kannel & McGee 1979a; Kannel & McGee 1979b; Ruderman & Haudenschild 1984).

1.1.2 (vi) Smoking as a risk factor

The US Department of Health and Human Services' 1983 report on the association between cigarette smoking and coronary heart disease (CHD) and other forms of cardiovascular disease estimated that up to 30% of deaths from CHD could be attributed to smoking (170 000 deaths in 1980). It also reported that smokers have a 70% higher CHD death rate than non-smokers, and those who smoke two or more packs per day have an almost 200% greater CHD mortality rate than non-smokers. A more recent study carried out by the Department of Health in 1994 (Report of the Cardiovascular Review Group of the Committee on Medical Aspects of Food Policy) has estimated that around one in eight deaths from CVD were attributable to smoking. A recent case-control study (INTERHEART study) carried out in 11119 cases and 13648 controls from 52 countries found an odds ratio of 2.87 for myocardial infarction (MI) among current smokers compared to those who had never smoked (Yusuf *et al.* 2004).

1.1.2 (vii) Socio-economic status as a risk factor

A low socio-economic status has been associated with risk of cardiovascular disease. Studies conducted in several countries have shown a continuous inverse relationship between socio-economic status, as measured by income, education, or professional grade, and cardiovascular morbidity and mortality (Brunner *et al.* 1997; Jacobsen & Thelle 1988; Matthews *et al.* 1989). However, the association can be partly explained by higher levels of cholesterol and blood pressure, and higher rates of smoking (Lynch *et al.* 1996; Jacobsen & Thelle 1988; Kadiri *et al.* 1999). A recent systematic review found that several risk factors including smoking, birth weight, adiposity (in young adults), height, and some aspects of diet, particularly fat and fibre consumption were all associated with lower socio-economic status (Batty & Leon 2002). Nevertheless, it is evident that early and later life deprivation through lower socio-economic status has an impact on health, in particular, risk of CVD mortality. Other studies, such as the Whitehall II study, found an independent association of lower socio-economic status and risk of CHD in civil servants (Marmot *et al.* 1997). In addition, the authors showed that low job control and workplace stress was also associated with socio-economic status.

1.1.2 (viii) Birth weight as a risk factor

Low birth weight and a disproportionate head size, length, and weight are thought to be markers of a lack of nutrients or oxygen during foetal growth. Several epidemiological studies have shown associations between a low birth weight and an increased rate in later life of coronary disease, as well as stroke, hypertension and non-insulin dependent diabetes (Frankel *et al.* 1996; Rich-Edwards *et al.* 1997; Leon *et al.* 1998; Eriksson *et al.* 2001; Barker 2004). These observations underpin the Barker hypothesis, that proposes that physiological differences programmed *in utero* by nutritional deprivation leads to a phenotype that increases the risk of CVD (Barker 2004). Although supported by experimental studies in animals, counter views have emerged that suggest the association may be explained by genotype(s) that predispose(s) both to low birth weight and later CVD risk (IJzerman *et al.* 2001), or that the association may be overestimated, as a recent systematic review suggests the presence of publication bias, confounding and inappropriate adjustment (Huxley *et al.* 2004). The association between birth weight and later CVD also appears to be largely modified by growth during life where individuals who were born small and became much heavier during childhood appear to have the highest risk of coronary disease (Eriksson 2005).

1.1.2 (ix) Risk scores for cardiovascular disease

Age, cholesterol, blood pressure, diabetes and smoking are all used in the estimation of risk of coronary disease in clinical practice. Risk scores such as the PROCAM or Framingham risk equations in which these variables are incorporated in multivariate risk models can be used to estimate 10 year risk of CHD in a given individual (Assmann *et al.* 2002; Anderson *et al.* 1991). There risk estimates are used to guide decisions to intervene to reduce risk factors in the primary prevention of clinical events. Although it has been shown that combining information on several risk factors increases the overall screening performance compared to the predictive utility of the individual risk factors in isolation, the degree to which this occurs is not as great as might be expected (Wald *et al.* 2005).

1.1.2 (x) Geographical and ancestral influences

The prevalence of established risk factors are subject to geographical and ancestral variation. Differences may exist in risk factors between groups of differing ancestral origin within their indigenous geographical location. In addition, the effects of migration may give rise to an increased exposure to environmental risk factors such that an interaction between genetic and environmental exposures takes place. Studies have been carried out comparing risk factor prevalence and CHD among groups of differing ancestry in different countries. A study carried out by LH Kuller showed that there are very large differences in levels of risk factors for CHD between migrant populations and those of similar ancestry living in the geographical region of

origin, for example, comparing African origin subjects living in Africa with those in America, or Japanese individuals living in Japan with those in Hawaii and California (Kuller 2004).

1.1.3 Genetics of coronary heart disease

A family history has also been found to be an independent risk factor for CHD in a number of studies (Boer *et al.* 1999; Li *et al.* 2000). However, both genetic factors and a shared environment can result in familial clustering of CHD risk. Prospective studies looking mainly at the role of family history in early-onset CHD in men have arrived at the same conclusions (Rissanen 1979b; Rissanen & Nikkila 1979; Rissanen 1985; Rissanen & Nikkila 1984; Rissanen & Nikkila 1979; Rissanen 1979a; Rissanen & Nikkila 1984; Nora *et al.* 1980; Shea *et al.* 1984; Friedlander *et al.* 1985; Barrett-Connor & Khaw 1984). Heritability is the proportion of the total variance in a continuous trait (or disease) that is attributable to genes (i.e. the genetic source of variance divided by the total phenotypic variance). If a trait is determined by an allele for which everyone is homozygous, then heritability will be estimated as zero, falsely deflating the genetic contribution to the trait. Conversely, in a sample in which exposure to an environmental determinant is universal, there will be no measured contribution to the total phenotypic variance, therefore falsely inflating the genetic component.

Studies have attempted to quantify the size of the genetic component to CHD susceptibility by comparing concordance for CHD between monozygotic and dizygotic twin pairs (Snowden *et al.* 1982; Marenberg *et al.* 1994; Friedlander *et al.* 1997). These studies have shown there is a higher degree of CHD risk among monozygotic twins (who share 100% of their genes) compared to dizygotic twins (who share 50% of their genes). This risk appears greater in younger twins, becoming less pronounced in older individuals, perhaps as exposure to environmental risk factors accumulates (Zdravkovic *et al.* 2002). However, the sibling recurrence risks (defined as the risk of disease in a sibling of a subject affected by the disorder divided by the prevalence in the population) are relatively low for CHD (λ_s of around 3) compared to other complex diseases, such as multiple sclerosis or Type I diabetes (λ_s of around 15), suggesting that the effect of individual genes may be very small indeed and correspondingly difficult to ascertain reliably (Sadovnick *et al.* 2000; Farrall *et al.* 2006).

Molecular studies examining plasma lipids and blood pressure levels (Williams *et al.* 1993; Livshits & Gerber 2001) have demonstrated familial aggregation of these traits that underlie CHD risk. Family studies have suggested that these factors are under a moderate degree of genetic control, with heritability estimates ranging from 0.35 to 0.5 for total cholesterol, 0.45 to 0.88 for LDL-cholesterol, and 0.37 to 0.66 for HDL-cholesterol (Dahlen *et al.* 1983; Chen *et al.* 1984; Friedlander *et al.* 1986; Austin *et al.* 1987; O'Connell *et al.* 1988). The observed variance is likely to be influenced by a number of genes. Studies by Sing &

Davignon have shown that as much as 16% of the genetic variance for LDL-cholesterol could be accounted for by common variants at the apolipoprotein (apo) E gene locus (Sing & Davignon 1985). Similarly, Deeb et al. found association between variants of the apolipoprotein A and B genes and cholesterol, triglycerides and HDL (Deeb *et al.* 1986). In addition, common variation in the gene for cholesteryl ester transfer protein (CETP) has been shown to be associated with differences in HDL- cholesterol concentration (Inazu *et al.* 1994). More recently, mutations have been found in the proprotein convertase subtilisin/kexin type 9 (PCSK9) gene, which is involved in the post-translational degradation of the LDL-receptors in the liver. Carriers of common variants in this gene have concentrations of LDL- cholesterol that are lower by around 40% compared to non-carriers (Austin *et al.* 2004; Cameron *et al.* 2006; Humphries *et al.* 2006).

Although CHD is a complex disease, rarer monogenic forms of the disorder exist, such as familial hypercholesterolaemia (FH), an autosomal dominant disorder. In most populations the frequency of homozygotes is 1 in a million and heterozygotes is 1 in 500 (Wyndham *et al.* 1987; Seftel *et al.* 1989). The disorder most often arises from mutations in the LDL-receptor (LDLR) but mutations in apoB also give rise to the same phenotype (Hobbs *et al.* 1992; Russell *et al.* 1989; Russell *et al.* 1986; Farese *et al.* 1992). The LDL-receptor is a cell surface transmembrane protein that mediates the uptake and lysosomal degradation of plasma LDL by the liver. Affected individuals have reduced receptor expression or function and therefore have elevated plasma levels of LDL, causing premature atherosclerosis, with clinical events in the second or third decades of life (Austin *et al.* 2004). Mutations in the apoB gene are mostly nonsense or frameshift mutations that interfere with the translation of a full-length apoB100 molecule and result in very poor binding of apoB100 to the cellular LDL receptor. This delays the clearance of LDL from the plasma and results in hypercholesterolaemia and premature atherosclerosis (Boren *et al.* 1998).

The discovery of these disorders helped to establish the causal link between elevated LDL- cholesterol concentration and CHD and contributed to the development of statins that have been shown to be efficacious in reducing CHD risk in a wide range of individuals, not just those with LDL-R mutations. There is intense interest therefore that understanding the genetic influences of traits linked to CHD will help to evaluate their causal relevance, and could lead to the development of new therapies to prevent CHD events (see section on Mendelian randomisation).

1.1.4 Pathogenesis of atherosclerosis and established risk factors

Atherosclerosis may begin in childhood (McGill *et al.* 2000) and have a long pre-clinical phase, but it is the advanced stages of this condition in later life that lead to clinical events. The lesions of atherosclerosis are not distributed in a random fashion in the vasculature. Haemodynamic factors interact with the activated vascular endothelium leading to lesions in coronary, carotid and other vessels. Fluid shear stresses generated by blood flow influence the phenotype of the endothelial cells by modulation of gene expression and regulation of the activity of flow-sensitive proteins. Atherosclerotic plaques characteristically occur in regions of branching and marked curvature at areas of geometric irregularity and where blood undergoes sudden changes in velocity and direction of flow (Velican & Velican 1985; Velican *et al.* 1986). Decreased shear stress and turbulence may promote atherogenesis at these important sites within the coronary arteries, the major branches of the thoracic and abdominal aorta, and the large conduit vessels of the lower extremities (Stone *et al.* 2003).

There have been three main theories on the pathogenesis of atherosclerosis (Hort 1994; Hort 2002). The first is the encrustation theory, developed by von Rokitansky, in which it was proposed that small thrombi consisting of platelets, fibrin and leukocytes cluster over sites of endothelial injury and interact with smooth muscle cell proliferation to form atherosclerotic plaques. The irritation theory, initiated by Virchow, proposed that proteins and lipids in the blood instigate an inflammatory reaction, causing cellular proliferation in the intima, giving rise to the atherosclerotic condition he named "chronic endoarteritis deformans". The third theory, as initiated by the experiments of Anitschkov and Chalatov in rabbits fed a cholesterol-enriched diet, is known as the lipid infiltration theory, and motivated a huge mass of studies devoted to plasma lipids and their cellular metabolism (Klimov & Nagornev 2002).

In more recent years, Ross proposed a "response to injury" hypothesis, which takes into account interactions among all of the cells found in the lesions of human atherosclerosis, and the cytokines and growth factors that can be formed by each of these cells. It suggests that lesions of atherosclerosis arise as a result of some form of injury to arterial endothelium, leading to alterations in endothelial cell-cell attachment or endothelial cell-connective tissue attachment and resulting in denudation of the endothelium. This, it was proposed, then allows platelets, monocytes and leukocytes to adhere to the wall at sites of injury, permitting plasma constituents to gain entry to the vessel wall and leads to smooth muscle cell proliferation. This theory was later modified by the concept that the endothelium only needs to be dysfunctional rather than denuded to initiate atherosclerosis (Ross *et al.* 1977; Ross 1990; Ross & Agius 1992; Ross 1997).

Injurious exposures might include increased blood pressure (Calver *et al.* 1992; Agmon *et al.* 2000), smoking (Celermajer *et al.* 1993) and increased cholesterol (Schwartz *et al.* 1989; Celermajer *et al.* 1996). Mechanistic studies have found an abnormality of basal nitric oxide (NO)-mediated dilation in patients with all of these risk factors, showing the NO system is impaired (Vita *et al.* 1990; Calver *et al.* 1992; Meredith *et al.* 1996; Kincer *et al.* 2002; Ambrose & Barua 2004). Elevated serum levels of LDL- cholesterol may also cause endothelial dysfunction by helping to provide oxidising radicals that reduce the bioavailability of NO, which is antithrombotic, anti-inflammatory and anti-atherosclerotic. NO is a freely diffusible gas generated by nitric oxide synthase (NOS) from L-arginine.

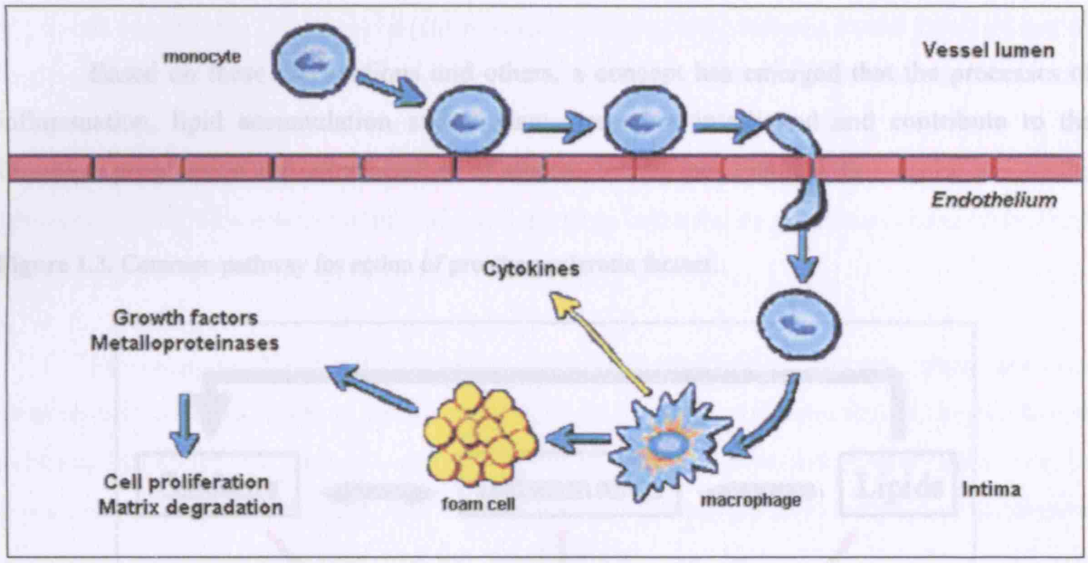
Reactive oxygen species (ROS) are thought by some to play an important role in much cardiovascular pathology involving inflammatory processes, including atherosclerosis. Several pathogenic effects of increased ROS production have been proposed. These include the oxidation of core lipids of lipoproteins and cell membranes, which may then modify apolipoproteins and other proteins (Griendling & Alexander 1997). ROS may be responsible for many aspects of vascular dysfunction and atherogenesis by modulating the expression of genes that influence the recruitment of circulating cells into the arterial intima, cell proliferation, and apoptosis. Exposure of cells to ROS induces a local inflammatory response and the release of cytokines and growth factors, such as tumour necrosis factor-α (TNF-α), interleukin-1β (IL-1β), and interferon-γ. These may then provide a positive feedback mechanism, for example by stimulating ROS-producing enzymes (Ross 1993). The precise mechanisms by which ROS can be generated in the vascular wall are the subject of much research. Potential sources include cytochrome enzymes, NADPH oxidases, xanthine oxidase and the mitochondrial electron transport chain. The physiological (as opposed to pathological) role of these enzymes is also a subject of interest (Ray & Shah 2005; Cave *et al.* 2006).

In the presence of oxidative stress LDL particles could become oxidized to form a lipoprotein species that may be particularly atherogenic. Several factors may influence the susceptibility of LDL to oxidation, including its size and composition, and the presence of endogenous antioxidant compounds, such as α-tocopherol. Individuals with type 2 diabetes or the metabolic syndrome have high levels of oxidative stress and this might contribute to the increased risk for cardiovascular events (Rosenson 2004). Oxidized LDL is capable of a wide range of toxic effects and cell and vessel wall dysfunctions that are characteristically and consistently associated with the development of atherosclerosis. These dysfunctions include impaired endothelium-dependent dilation and paradoxical vasoconstriction, and are the result of direct inactivation of NO by the excess production of free radicals, reduced transcription of NOS mRNA, and post-transcriptional destabilisation of mRNA (Bhakdi 2003). The decrease in the availability of NO also is associated with increased platelet adhesion, increased plasminogen

activator inhibitor, decreased plasminogen activator, increased tissue factor, decreased thrombomodulin, and alterations in heparan sulfate proteoglycans. Oxidised LDL may also promote atherosclerosis through the transformation of macrophages into foam cells that results from the uptake of large amounts of oxidised LDL via scavenger receptors (Napoli *et al.* 2001). Despite these extensive plausible mechanistic studies, randomised controlled trials of anti-oxidant vitamins (vitamin c and vitamin e) have not indicated benefits of such interventions in preventing CHD events, casting doubt on the role of ROS in coronary disease development (Roncaglioni 2001; Heart Protection Study Collaborative 2002; Lawlor *et al.* 2004).

The endothelial injury leads to a variety of responses by the endothelium including expression of a variety of cytokines that lead to local and systemic effects. In addition, the antithrombotic surface of the endothelium changes to a prothrombotic one. Cytokines chemotactic for monocytes are released, recruiting these cells to the growing lesion (see Figure 1.2). The injured endothelium expresses adhesion molecules to which these monocytes subsequently bind to, traversing the endothelial cell layer. Here they ingest modified LDL, becoming lipid-laden macrophages that secrete growth factors for smooth muscle cells (SMCs) (Ikeda *et al.* 1998). The cholesterol engorged macrophages are then transformed into foam cells. The monocytes can ingest enough LDL to rupture releasing free cholesterol, cytokines, and procoagulants into the surrounding environment, forming a fatty streak (Fan & Watanabe 2003). Furthermore, oxidized LDL activates inflammatory processes at the level of gene transcription by up-regulation of nuclear factor kappa-B (NFκB), which result in further expression of adhesion molecules by endothelial cells, and recruitment of additional monocytes (Libby *et al.* 2000).

Figure 1.2. Proposed initial stages of atherogenesis. Monocytes bind to the endothelium and pass through into the intima, where they ingest modified LDL and become macrophages, which are then transformed into foam cells.



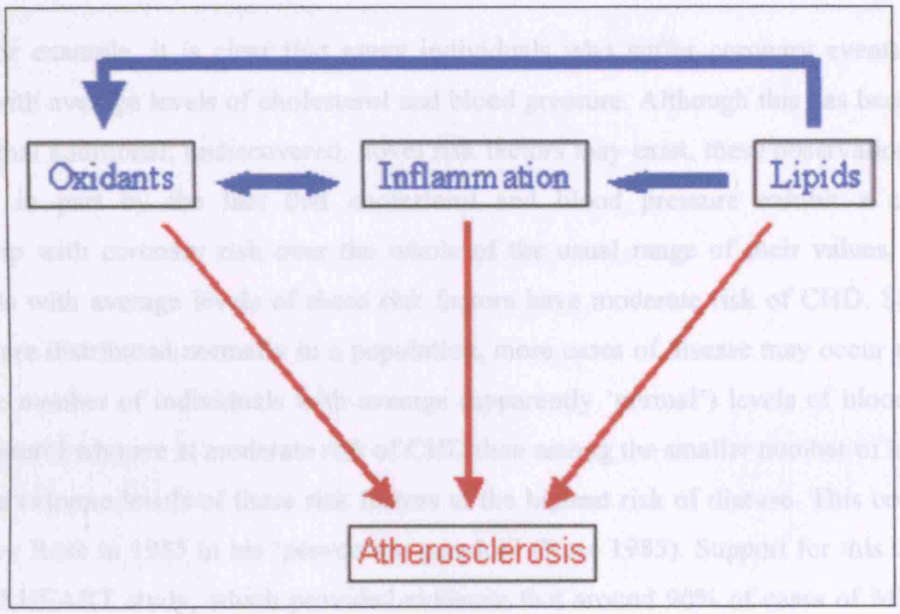
It is thought that progressive lipid accumulation and the migration and proliferation of SMCs may result in the progression of the fatty streak to form a fibrous plaque. Platelet-derived growth factor, insulin-like growth factor, transforming growth factors alpha and beta, thrombin, and angiotensin II are potent mitogens that are produced by activated platelets, macrophages, and dysfunctional endothelial cells and these are all expressed in proliferating plaques. The relative deficiency of endothelium-derived nitric oxide further potentiates this proliferative stage of plaque maturation (Libby 1997). The SMCs at the site are thought to be responsible for the deposition of extracellular connective tissue matrix and form a fibrous cap that overlies a core of lipid-laden foam cells, extracellular lipid, and necrotic cellular debris. Growth of the fibrous plaque is likely to result in vascular remodelling, progressive luminal narrowing, blood-flow abnormalities, and compromised oxygen supply to the target organ (Gordon *et al.* 2001).

Denudation of the overlying endothelium or rupture of the protective fibrous cap due to weakening may result in exposure of the thrombogenic contents of the core of the plaque to the circulating blood. This exposure constitutes an advanced or complicated lesion (Libby 1997). Inflammatory cells localise to the shoulder region of the vulnerable plaque; T lymphocytes elaborate interferon gamma, an important cytokine that impairs vascular smooth muscle cell proliferation and collagen synthesis. Furthermore, activated macrophages produce matrix metalloproteinases that degrade collagen and may weaken the fibrous cap, rendering the plaque

more vulnerable to rupture. A plaque rupture may result in thrombus formation, partial or complete occlusion of the blood vessel through the initiation of the clotting cascade, or progression of the atherosclerotic lesion due to organisation of the thrombus and incorporation within the plaque.

Based on these observations and others, a concept has emerged that the processes of inflammation, lipid accumulation and oxidant stress are interlinked and contribute to the initiation, progression and complications of atherosclerosis (see Figure 1.3).

Figure 1.3. Common pathway for action of proatherosclerotic factors.



1.2 Inflammation and Atherosclerosis

1.2.1 A role for novel risk factors in atherosclerosis

A central role for lipids, blood pressure, smoking and diabetes in the development of cardiovascular disease has been known for some time. In the case of cholesterol, much has been discovered about the likely mechanisms through which LDL- cholesterol contributes to the development of atherosclerotic lesions, through the experimental studies described in the previous section. However, clinical observations have led to the proposal that conventional risk factors may not account for all cases of CHD (Braunwald 1997).

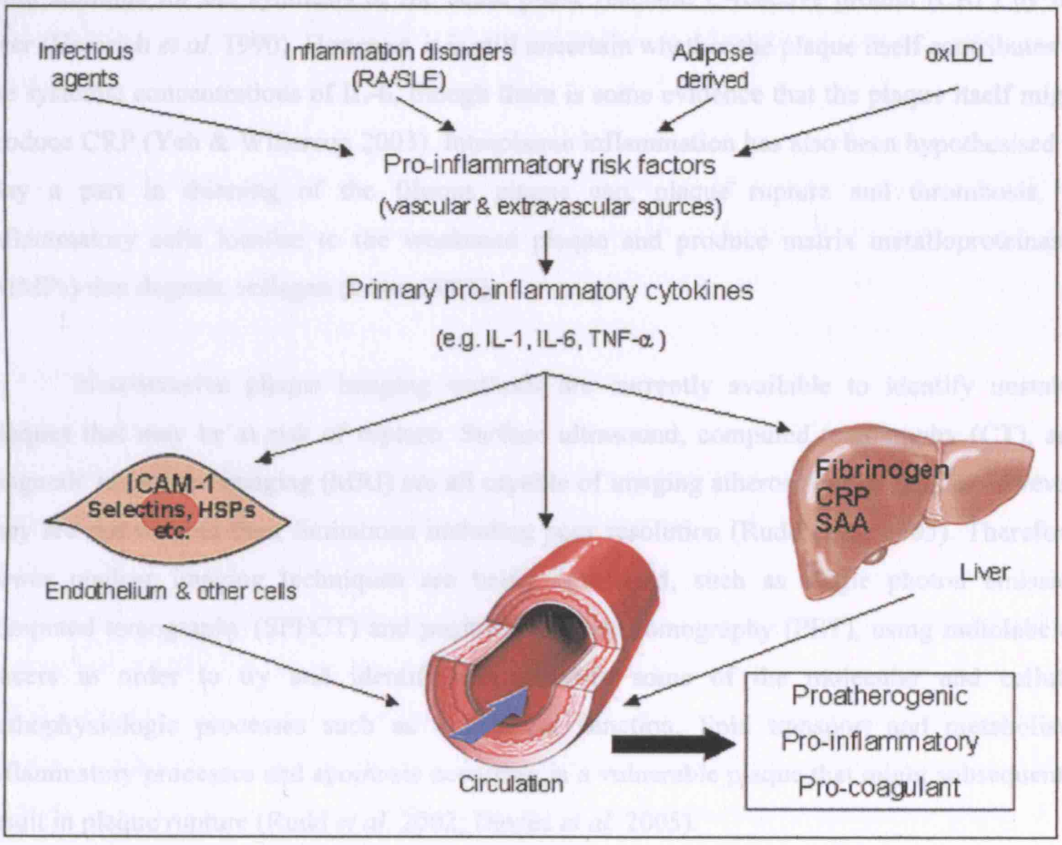
For example, it is clear that many individuals who suffer coronary events are non-smokers with average levels of cholesterol and blood pressure. Although this has been taken as evidence that additional, undiscovered, novel risk factors may exist, these observations may be explained in part by the fact that cholesterol and blood pressure exhibit a continuous relationship with coronary risk over the whole of the usual range of their values, such that individuals with average levels of these risk factors have moderate risk of CHD. Since these variables are distributed normally in a population, more cases of disease may occur among the very large number of individuals with average (apparently 'normal') levels of blood pressure and cholesterol who are at moderate risk of CHD than among the smaller number of individuals with more extreme levels of these risk factors at the highest risk of disease. This concept was outlined by Rose in 1985 in his 'prevention paradox' (Rose 1985). Support for this came from the INTERHEART study, which provided evidence that around 90% of cases of MI could be explained by smoking, raised ApoB/ApoA1 ratio, high blood pressure, diabetes, obesity, psychosocial factors, fruit and vegetable consumption, alcohol consumption, and physical exercise, although, because of the case-control design of this study, this might be an over-estimate (Yusuf *et al.* 2004). Nevertheless, there is still substantial interest in and evidence for a role for so-called "novel" pathways and mechanisms in coronary disease, notable among which is inflammation, for which substantial evidence for a role in CHD now exists.

1.2.2 Role of Inflammation in the development of CHD

Over the past decade much laboratory evidence has emerged that suggests that inflammatory processes within the atherosclerotic plaque itself are involved in the initiation and evolution of the atherosclerotic lesion, as well as contributing to the ultimate development of acute ischaemic syndromes (Ross 1999). Additionally, inflammatory processes remote from the plaque are also thought by some to modulate the progression and complications of atherosclerosis through generation of a systemic inflammatory response. These processes could

include infections in the mouth (periodontal disease), infections in the gut due to *Helicobacter pylori*, as well as inflammation in the joints as seen in arthritis. Expression and release of inflammatory cytokines such as IL-6 from adipose tissue that are upregulated in obesity may also contribute and it has been proposed they might contribute to the link between obesity, metabolic syndrome and CHD (Juge-Aubry *et al.* 2003) (see Figure 1.4). The evidence for these links will now be discussed in more detail.

Figure 1.4. Hypothesised link between infection, non-infectious inflammatory disorders and cardiovascular disease.



1.2.3 Inflammation within atherosclerosis

Studies of plaque morphology, cell population and gene expression have led to the proposal that when the endothelium becomes injured or inflamed, it becomes activated and expresses adhesion molecules such as intercellular adhesion molecule-1 (ICAM-1) that may facilitate attachment of inflammatory cells to the endothelium (Libby *et al.* 2002). These cells then migrate into the subendothelial space, attracted by monocyte chemoattractant protein-1 (MCP-1) (Gu *et al.* 1998). Here the leucocytes contribute to the local inflammatory response,

with macrophages expressing receptors for modified lipoproteins, and macrophage colony stimulating factor combining with MCP-1 to augment the differentiation of blood-derived monocytes into macrophage foam cells (Qiao *et al.* 1997). T-cell activation leads to expression of interferon-γ and lymphotoxin, which could further amplify the pro-inflammatory state.

Macrophages, endothelial cells and SMCs produce the cytokine TNF-α (Warner & Libby 1989; Barath *et al.* 1990), which along with interferon-γ and IL-1, is known to stimulate smooth muscle cell production of IL-6 (Sanceau *et al.* 1995; Ng *et al.* 1994). IL-6 gene transcripts are expressed in human atheroma (Seino *et al.* 1994; Rus *et al.* 1996) and IL-6 is the main stimulus for the synthesis of the acute phase reactant, C-reactive protein (CRP) by the liver (Heinrich *et al.* 1990). However, it is still uncertain whether the plaque itself contributes to the systemic concentrations of IL-6, though there is some evidence that the plaque itself might produce CRP (Yeh & Willerson 2003). Intraplaque inflammation has also been hypothesised to play a part in thinning of the fibrous plaque cap, plaque rupture and thrombosis, as inflammatory cells localise to the weakened plaque and produce matrix metalloproteinases (MMPs) that degrade collagen (Lowe 2001).

Non-invasive plaque imaging methods are currently available to identify unstable plaques that may be at risk of rupture. Surface ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) are all capable of imaging atherosclerotic plaque, however, they are not without their limitations including poor resolution (Rudd *et al.* 2005). Therefore, newer nuclear imaging techniques are being developed, such as single photon emission computed tomography (SPECT) and positron emission tomography (PET), using radiolabeled tracers in order to try and identify and quantify some of the molecular and cellular pathophysiologic processes such as endothelial function, lipid transport and metabolism, inflammatory processes and apoptosis occurring in a vulnerable plaque that might subsequently result in plaque rupture (Rudd *et al.* 2002; Davies *et al.* 2005).

1.2.4 Evidence for a role for inflammation in atherogenesis from animal studies

Although inflammation within the plaque is thought to enhance progression of atherosclerosis, even inflammatory processes at distant sites may play a part through generation of a systemic inflammatory response. Studies in experimental animal models have shown that infectious agents can accelerate the development and severity of atherosclerosis. Studies carried out in chickens infected with Marek's disease virus (MDV), a type of herpesvirus, have shown that infected chickens on a cholesterol-supplemented diet developed atherosclerosis that is often occlusive. Infected chickens on a normal diet also developed atherosclerosis, but the lesions were less fatty, suggesting additive effects of infection and cholesterol. All of the infected

chickens had lesions that resemble those seen in humans. In contrast, none of the uninfected controls had visible atherosclerosis (Fabricant *et al.* 1978).

Infection with other viruses, such as murine herpes virus 68 (MHV-68) or cytomegalovirus (CMV) appears to accelerate atherosclerosis in Apo-E knockout mice (Alber *et al.* 2000; Hsich *et al.* 2001; Hsich *et al.* 2001). Apo-E knockout mice on a normal diet had high cholesterol concentrations and spontaneously developed atheroma that resembled the human disease. Infection with MHV-68 has been shown to enhance atherosclerosis without altering lesion histology. Infection with CMV leads to increased atherosclerotic lesion size and number in Apo-E knockout mice (Hsich *et al.* 2001; Vliegen *et al.* 2004; Vliegen *et al.* 2004) and has also recently been shown to increase aortic expression of atherogenic genes (Burnett *et al.* 2004).

Early animal studies on the cellular origin and development of radiation-induced atheromatous plaques in the carotid artery of hypercholesterolemic rabbits have shown that invading monocytes transform into lipid-laden macrophages and become foam cells, and eventually result in plaques (Vos *et al.* 1983). The results suggest that intra-plaque inflammation promotes lipid accumulation, promoting atherosclerosis.

Other infectious agents, such as *Chlamydia pneumoniae*, have also been studied in relation to potential effects on atherosclerosis (Grayston *et al.* 1997; Grayston *et al.* 1997). New Zealand White rabbits fed a normal diet and inoculated with *C. pneumoniae* have shown evidence of bronchiolitis and pneumonia (Moazed *et al.* 1996; Fong *et al.* 1997) and from histological studies, early and intermediate lesions of atherosclerosis were seen (Fong *et al.* 1997). Infection has also been shown to accelerate intimal thickening in rabbits given a modestly cholesterol-enhanced diet that can be prevented by weekly antibiotic treatment with azithromycin (Muhlestein *et al.* 1998), although delayed treatment is ineffective (Fong *et al.* 1999). In hypertensive Tg53 rats, *C. pneumoniae* infection induced acceleration of foam cell formation from peritoneal macrophages, which is hyperlipidemia-dependent, and coronary plaques and left atrial thrombi can be seen (Herrera *et al.* 2003).

Periodontal diseases are localized chronic inflammatory conditions of the gingiva and underlying bone and connective tissues induced by bacteria and bacterial products of dental plaque, such as invasion by the periodontal pathogen *Porphyromonas gingivalis*, a gram-negative anaerobe that colonises the subgingival region in the mouth and can modulate eukaryotic cell signal transduction pathways, thus altering the production of cytokines and other components of the immune response (Deshpande *et al.* 1998; Lamont & Jenkinson 1998). Studies in the Apo-E knockout model have supported a causal link between periodontal

infection and atheroma by demonstrating that inoculation with *P. gingivalis* accelerates atherosclerosis (Li *et al.* 2002; Lalla *et al.* 2003; Gibson, III *et al.* 2004). This mechanism might be because a virulence factor(s) such as lipopolysaccharide (LPS) enters the circulation to initiate or promote foam cell formation in macrophages in the arterial wall or through initiation of a systemic inflammatory response (Qi *et al.* 2003).

Atherosclerosis is accompanied by a local immune response in the plaque, but its role in the pathogenesis of the disease is still unclear. To try and elucidate its role, studies have been carried out where atherosclerosis-prone Apo-E knockout mice were crossed with immunodeficient *scid/scid* mice. The offspring had significantly fewer atherosclerotic lesions, although the number of lesions could be increased by the reconstitution of immune function by the transfer of $CD4^+$ T-cells from Apo-E knockout donors to immunodeficient recipients (Zhou *et al.* 2000; Hansson 2001). The recipient mice also had increased levels of interferon-γ, which has been shown to accelerate transplant vascular lesions in arteries xenografted into *scid/scid* mice (Tellides *et al.* 2000).

There appears to be good evidence for the role of inflammatory processes in atherosclerosis in animal models. One reason for the large amount of evidence is the ease of modifying animal systems and creating knockouts to probe specific effects of single pathways or molecules while keeping all other factors constant. However, it is more difficult to ascertain the role of inflammatory processes in atherosclerosis in humans.

1.2.5 Human infections and atherosclerosis

Acute infections such as respiratory or urinary infections or sepsis, have been associated with an increased risk of acute athero-thrombotic events such as stroke and MI in the few days or weeks that follow infection. Chronic low-grade infections with organisms such as cytomegalovirus, *C. pneumoniae*, *H. pylori* and periodontal disease, have been associated with risk of coronary events years later. The robustness of the evidence in these two areas is somewhat different.

Several studies have demonstrated that acute infections may instigate an inflammatory response in the vessel wall, drive an immune response or induce autoimmunity against cell wall components, all of which may accelerate atherosclerosis. Temporary increases in risk of CVD have been reported following a wide range of inflammatory stimuli, including surgery (Aitkenhead 1993; Mamode *et al.* 1995), respiratory and urinary infections (Grimes *et al.* 2000; Smeeth *et al.* 2004) and sepsis (Leelawattana *et al.* 2003). While all of these studies are observational and could therefore be prone to confounding or certain biases, the study by

Smeeth et al. utilising routinely collected data from the UK General Practice Research Database may be less susceptible to such errors. This is because a case-series design was used, with each subject as their own control with the hypothesis tested that the rate of CHD events or stroke within a month of respiratory or urinary infection was higher than at any other time (Smeeth *et al.* 2004).

In further support of the important role of inflammation in the pathogenesis of the complications of atherosclerosis, patients with elevated levels of IL-6 and IL-1 receptor antagonist (IL-1Ra) at 48 h after admission for acute coronary syndromes were at increased risk of a complicated in-hospital course, and their levels can be used to predict outcome (Biasucci *et al.* 1999c). Following acute infection, cytokines such as IL-6 induced an acute phase response that was associated with an increased risk of thrombotic events, perhaps through increasing expression of fibrinogen, plasminogen activator inhibitor-1 (PAI-1) and tissue factor, and by increasing platelet expression of CD40 (Woodhouse *et al.* 1997; Roumen-Klappe *et al.* 2002).

Much attention has also focused on infectious sources as potential instigators of chronic low-grade inflammation that may increase the risk of atherosclerosis. Epidemiological studies have shown an association between atherosclerosis and antibodies against CMV, a member of the herpesvirus family (Melnick *et al.* 1995). However, concerns have been raised about the interpretation of these observational studies owing to the small sample sizes, potential publication bias, potential for confounding from other risk factors such as smoking, and the analysis of only a few patients with classical atheromatous lesions in individual studies (Danesh *et al.* 1997). The bacterial agents that have been linked to atherosclerosis in recent years include *H. pylori* and *C. pneumoniae*. Evidence for the *H. pylori* association is primarily based on antibody studies (Danesh *et al.* 1997), as this microorganism has not been detected in any atheromatous plaques, although it has been shown to induce a distant inflammatory response from the gut.

Evidence for an association of *C. pneumoniae* with atherosclerosis may also be prone to publication bias and confounding by established risk factors such as age, smoking, hypertension, and hyperlipidemia (Danesh *et al.* 1997; Grayston *et al.* 1997). A recent meta-analysis of 15 prospective studies found an odds ratio of 1.15 (0.97 to 1.36) for coronary disease between the top and bottom tertiles for serum *C. pneumoniae* IgG titres, which was attenuated after adjustment for measured confounders (Danesh *et al.* 2000b). However, since all potential confounders may not have been measured, this odds ratio may be attenuated further, suggesting the effect size, if any, may be very small.

One method of overcoming the problem of confounding to test the true causal relevance of infectious exposures is to carry out randomised controlled trials (RCTs), where subjects are randomised to an intervention to test the infectious agent or a control. The randomisation balances the known and unknown confounders between the groups so that any difference in outcome should be through the treatment in question. A number of trials have been conducted evaluating the effect of antibiotics effective against *C. pneumoniae* on CHD events. Early small trials of azithromycin, such as the azithromycin trial in survivors of MI (Gupta *et al.* 1997) and the Azithromycin in Coronary Artery Disease: Elimination of Myocardial Infection with Chlamydia (ACADEMIC) study (Anderson & Muhlestein 2000), have shown that antibiotic treatment reduces the *C. pneumoniae* titre count, and produced some reduction in clinical events. However, larger studies, such as the WIZARD trial, with longer periods of azithromycin antibiotic treatment, extended follow-up times and sufficient numbers of clinical events to provide a more reliable test of the benefits of antibiotic treatment, have been negative (O'Connor *et al.* 2003). A recent combined analysis of three large trials, the WIZARD, ACES and PROVE-IT trials, found that the overall percentages for CHD were 19% in the treatment group and 20% in the control group, which was not statistically significant (Danesh 2005). This suggests that the positive associations between *C. pneumoniae* seropositivity and CHD risk seen in observational studies could be due to confounding, calling the role of *C. pneumoniae* in atherosclerosis into question.

Chronic periodontal disease infections have also been associated with an increased risk of CVD. A prospective study carried out in 1993 suggested evidence of an association between dental health and CHD, as those with periodontitis had a 25% increased risk of CHD relative to those with minimal periodontal disease, and this was especially seen in males younger than 50 years (relative risk of 1.72) (DeStefano *et al.* 1993). However, this study could also be subject to confounding, as periodontal disease severity is strongly associated with smoking, an established risk factor for CHD (Stoltenberg *et al.* 1993; Turnbull 1995; Airila-Mansson *et al.* 2004). Randomised trials of the effect of periodontal disease treatment on vascular function and clinical events are currently underway and will allow the causal relevance of periodontal disease to be better evaluated.

1.2.6 Human inflammatory diseases and atherosclerosis

In addition to low-grade chronic infections, non-infectious inflammatory processes have also been postulated to be a stimulus for atherosclerosis. These include autoimmune diseases such as systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA), which have been linked with an increased risk of myocardial infarction. However, these associations may also be affected by confounding by conventional risk factors for cardiovascular disease.

Blood pressure levels and lipid profiles in SLE patients are often abnormal compared to the normal population due to exacerbations of disease and renal involvement, as well as the effect of treatment with steroids, therefore, it is not known whether SLE increases CHD risk through inflammation or through increased lipid levels (Leong *et al.* 1994). Later studies looking at lipid metabolism have found that antibodies to oxidised LDL are raised in the sera of SLE patients, and oxidised LDL containing immune-complexes are also elevated, however, it is not yet clear whether the increased prevalence of antibodies is pathogenic (George *et al.* 1999).

Abnormal lipid profiles are also seen in RA and increased concentrations of oxidised LDL have been found in the sera and synovial fluid from patients that increase the presence of foam cells and fatty streaks (Winyard *et al.* 1993). Therefore, again it is not known whether RA works through inflammation to increase CHD risk, or whether part of the effect is through classical risk factors. Circulating cytokines that are involved in the acute phase pathway are also markedly exaggerated in RA, as chronic cytokine release can also occur from inflamed joints. Due to the pleiotropic nature of these cytokines, they can also mediate numerous metabolic effects, including alterations in lipids and peripheral insulin resistance, and therefore elevated cytokine levels could influence atherosclerosis development through an inflammatory pathway or via more established risk factors (Sattar *et al.* 2003). Another possible explanation for the link is that RA, SLE and CHD might share underlying genetic susceptibility, perhaps through genes that regulate immune responses.

1.2.7 Circulating markers of inflammation and atherosclerosis

One of the physiological responses to inflammation is the acute phase response, which involves alterations in metabolism and gene regulation in the liver. During the acute phase response following an inflammatory stimulus, the liver synthesizes a range of plasma proteins known as acute phase reactants, whose concentration in the plasma changes. Most acute phase reactants are synthesised by hepatocytes, under the control of several cytokines such as IL-6 and IL-1, although some are produced by other cell types, including monocytes, endothelial cells, fibroblasts and adipocytes (Tracy 1999). Most acute phase reactants are positive acute phase reactants in that their levels can increase from several-fold to 1000-fold over normal levels during infection or inflammation. These include fibrinogen, CRP and serum amyloid A (SAA). Negative acute phase reactants include albumin, transthyretin, and transferrin, whose plasma levels can fall by upto 20% following tissue injury (Danesh *et al.* 1998; Robert 1999). Concentrations of many inflammatory cytokines and acute phase reactants have been associated with the risk of subsequent CHD events after an intial infective or inflammatory stimulus, or in overtly healthy subjects, with the risk of CHD years later.

Fibrinogen is an acute phase protein that can affect haemostasis, blood viscosity and leukocyte adhesion. It is elevated in healthy subjects with inherited predisposition for myocardial infarction (Margaglione *et al.* 2000). In addition, a recent meta-analysis of individual participant data showed association between fibrinogen concentration and risk of CHD and stroke (Fibrinogen studies Collaboration 2005). IL-6 concentrations have been measured in large studies such as the Physicians Health study (Ridker *et al.* 2000b) and have been found to be associated with future cardiovascular events. IL-6 and IL-1Ra levels are also associated with the risk of in-hospital complications following unstable angina (Biasucci *et al.* 1999c). TNF-α is a multifunctional circulating cytokine derived from endothelial and smooth muscle cells, as well as macrophages. It plays a major role in the cytokine cascade as it stimulates the synthesis of the other cytokines. It has been shown that recurrent ischaemic events are more common following an acute MI in patients with elevated TNF-α concentrations than those with concentrations lower than 4.17pg/mL (Ridker *et al.* 2000c). Numerous studies have been conducted on CRP showing that its circulating concentrations are consistently associated with risk of future cardiovascular events in apparently healthy individuals (as discussed in the next section) (Ridker *et al.* 1998a; Koenig *et al.* 1999; Danesh *et al.* 2000a; Ridker *et al.* 2003; Danesh *et al.* 2004). SAA levels have been demonstrated to be significantly higher in patients with severe atherosclerosis than in normal patients (Mezaki *et al.* 2003). Cellular adhesion molecules such as VCAM-1 and ICAM-1 have been detected in plasma and have been reported to be indicative of the expression of membrane bound adhesion molecules (Schmidt *et al.* 1995), and are correlated with future events in individuals with CAD (Blankenberg *et al.* 2001; Blankenberg *et al.* 2003).

All the associations seen between these inflammatory biomarkers and risk of coronary disease are consistent, but may be subject to confounding and reverse causality (as discussed in the section on Mendelian randomisation). Adjustment for confounding variables has shown attenuation of association in many cases, suggesting that the association may not be as strong as initially thought (Danesh *et al.* 1998; Danesh *et al.* 2000a). Moreover, concerns about residual confounding leaves considerable uncertainty as to the causal relevance of these factors and whether therapeutic intervention to lower their levels would reduce CHD events. This issue is the subject of several of the studies described later in this thesis. Since CRP is the most studied of these factors, and the subject of this thesis, the subsequent section provides a detailed background to the synthesis, regulation and known biology of CRP and the evidence linking CRP to coronary disease available at the time this thesis was started. This information was used to form the hypotheses, delineate the aims and design the experiments reported in this thesis.

1.3 C-Reactive Protein (CRP)

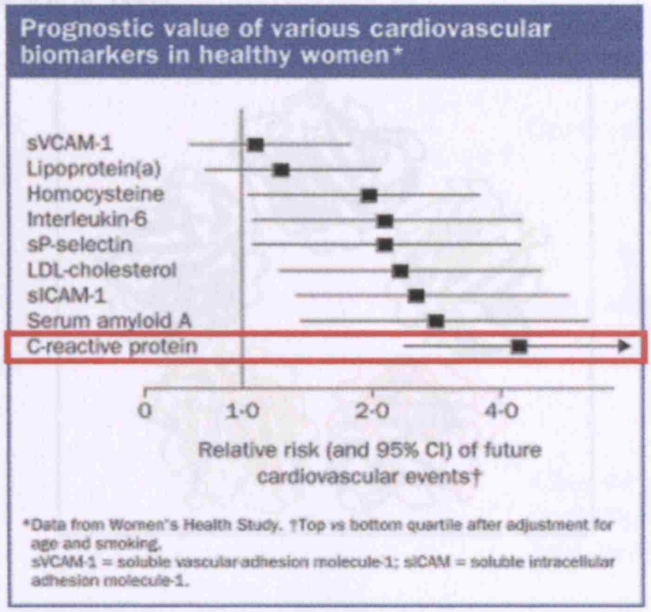
1.3.1 Associations of CRP with cardiovascular events

In keeping with the intense interest in the link between inflammation and vascular disease, there has recently been substantial interest in CRP, as a marker for future risk of coronary disease and as a potential mediator in atherogenesis. CRP is consistently associated with the risk of recurrent coronary events in individuals with an acute coronary syndrome (unstable angina, MI), in whom CRP concentrations rise to between 10-100 times their baseline values. Higher than average CRP concentrations are also associated with an increased risk of coronary events occurring in healthy individuals many years later (see Figure 1.5). Several prospective observational studies have reported this association and while the strength of the association has attenuated in larger studies with better adjustment for confounding, the relationship between CRP and coronary disease is consistently seen (Ridker & Haughie 1998; Danesh *et al.* 1998; Danesh *et al.* 2000a; Ridker 2001b; Ridker *et al.* 2002; Blake *et al.* 2003).

One of the reasons for the focus on CRP, as opposed to the many other potential markers of inflammation has been the development of simple, reliable and sensitive assays that have allowed CRP concentrations to be quantified even at the low concentrations seen in healthy individuals. Aside from the acute peaks of CRP seen in acute infections or inflammation, the concentration of CRP measured in healthy individuals seems as stable from year to year as concentrations of cholesterol (Hirschfield & Pepys 2003).

Although CRP shows consistent association with future coronary risk, the explanation for this relationship is incompletely understood. There is the exciting possibility that CRP may be causal in coronary disease. There is also the potential for its application as a tool to identify subjects of increased risk of coronary disease. The aim of this thesis is to address these issues. In order to place the new findings in context, the subsequent sections provide a detailed review of the known regulation and function of CRP.

Figure 1.5. Cardiovascular markers and relative risk of future cardiovascular events (Ridker 2001b).

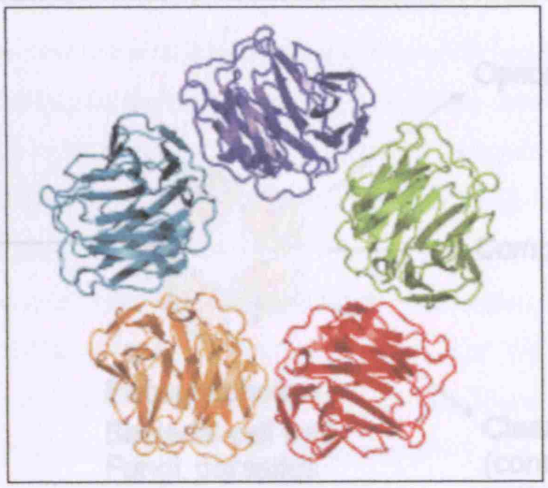


Data from the Women's Health Study shows that CRP has a relative risk of more than 4.0 for future cardiovascular events after adjustment for age and smoking.

1.3.2 CRP as an acute phase reactant

C-reactive protein was identified in 1930 by William S. Tillet and Thomas Francis at the Rockefeller Institute for Medical Research as an acute-phase protein (Tillet & Francis 1930). Acute-phase proteins are circulating proteins synthesised by the liver whose plasma concentrations increase or decrease substantially following an inflammatory stimulus. CRP is a member of the pentraxin family of proteins, which also includes serum amyloid P component (SAP), neuronal pentraxin and PTX3 (Pentraxin 3). CRP comprises five identical subunits (protomers) linked by noncovalent binding (see Figure 1.6). Each subunit comprises 206 amino acids in a single polypeptide chain with a total molecular weight of approximately 23 000 Daltons. The two cysteine residues at positions 36 and 78 form a disulfide bond (Volanakis 2001). CRP concentrations range from 0.1-5mg/L in healthy individuals and can increase to up to 1000mg/L within 24-48 hours following infection and inflammation.

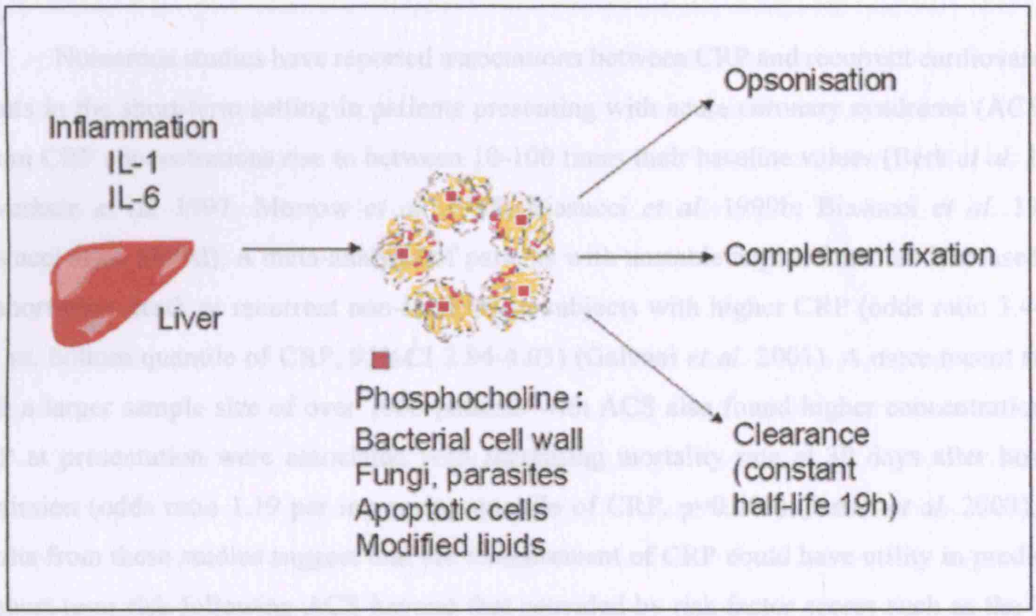
Figure 1.6. Three dimensional pentameric structure of C-reactive protein viewed down the 5-fold symmetry axis (Thompson *et al.* 1999).



1.3.3 Established functions of CRP

In man, CRP is a major acute-phase plasma protein displaying rapid and pronounced rise in concentration in response to infection or tissue injury. In rodents, CRP is not an acute-phase reactant, and is expressed at constitutively high levels in rats and very low levels in mice, which do not increase prominently following an inflammatory stimulus (Pepys *et al.* 1979; Szalai 2002b; Szalai 2002a). Nevertheless, mice expressing an acute-phase inducible human CRP transgene make a good model for studying the *in vivo* function of the protein since human CRP can fix rodent complement so any effects of CRP will be due to the human transgene and not murine CRP (Ciliberto *et al.* 1987). Human CRP has Ca^{2+}-dependent binding specificity for phosphocholine (PCh), a constituent of many bacterial and fungal polysaccharides and of most biological cell membranes (Volanakis & Kaplan 1971). Ligand-complexed CRP is recognized by C1q, a subunit of the C1 enzyme complex, and activates the classical pathway of human complement (Kaplan & Volanakis 1974; Volanakis 1982a; Volanakis 1982b). Its ability to recognize pathogens and to mediate their elimination by recruiting the complement system and phagocytic cells makes CRP an important constituent of the first line of innate host defence (see Figure 1.7).

Figure 1.7. Physiological role of CRP in opsonisation, complement fixation and clearance.



Although CRP concentrations increase during many infections and inflammatory disorders, such as systemic lupus erythematosus (SLE), rises in CRP during flares of disease activity are small although the CRP response to infection in patients with SLE is preserved (Pepys & Hirschfield 2003). It is interesting to note that SLE is associated with increased risk of cardiovascular events, however CRP levels do not rise with increasing autoimmune disease activity, perhaps suggesting that in SLE at least, the role of CRP in atherosclerosis is not a causal one. Indeed, a study looking at polymorphisms in the CRP gene and SLE has found that a polymorphism associated with reduced basal CRP was also associated with the development of SLE, suggesting high CRP levels may protect from the development of SLE (Russell *et al.* 2003).

Support for a protective role for the acute phase elevation of CRP in defence from infection comes from animal studies using transgenic mice expressing rabbit CRP. When compared to mice not overexpressing CRP, the transgenic mice were relatively protected from lethal bacterial endotoxemia (Xia & Samols 1997). Transgenic mice expressing human CRP are also protected against lethal infection by *Streptococcus pneumoniae* and *Salmonella enterica* serovar Typhimurium, showing that that CRP plays an important role in vivo in host defense during the early stages of infection. In addition, enhancement of the host humoral immune response was seen, suggesting that CRP may also contribute indirectly to host defense during later stages of infection (Szalai *et al.* 2000b). As murine CRP is not an acute phase protein, the effects seen in these transgenic mice must be solely due to the human homologue.

1.3.4 CRP and acute cardiovascular events

Numerous studies have reported associations between CRP and recurrent cardiovascular events in the short-term setting in patients presenting with acute coronary syndrome (ACS), in whom CRP concentrations rise to between 10-100 times their baseline values (Berk *et al.* 1990; Haverkate *et al.* 1997; Morrow *et al.* 1998; Biasucci *et al.* 1999b; Biasucci *et al.* 1999a; Biasucci *et al.* 1999d). A meta-analysis of patients with unstable angina found an increased risk of short-term death or recurrent non-fatal MI in subjects with higher CRP (odds ratio 3.44 for top vs. bottom quantile of CRP, 95%CI 2.94-4.03) (Galvani *et al.* 2001). A more recent study, with a larger sample size of over 7000 patients with ACS also found higher concentrations of CRP at presentation were associated with increasing mortality rate at 30 days after hospital admission (odds ratio 1.19 per increasing quartile of CRP, p=0.006) (James *et al.* 2003). The results from these studies suggest that the measurement of CRP could have utility in prediction of short-term risk following ACS beyond that provided by risk factor scores such as the TIMI risk score used in these settings that is based on the presence or absence of several clinical parameters, or sensitive assays of myocardial injury such as troponin assays (Foussas *et al.* 2005).

1.3.5 CRP and the longer-term process of atherosclerosis

A series of prospective observational studies have shown that even a mild elevation of baseline levels of CRP within the range normally encountered in health is associated with higher long-term risk for future coronary events years later, similar to the associations of cholesterol or blood pressure and coronary risk (Ridker & Haughie 1998; Danesh *et al.* 1998; Danesh *et al.* 2000a; Ridker *et al.* 2002). The associations appear to be graded continuous relationships over the whole range of CRP concentrations without a threshold. Such associations may indicate a potentially causal role for CRP in atherosclerosis, and therefore that CRP could provide a legitimate target for therapy, much like cholesterol or blood pressure. However, observational studies have their limitations (as discussed in Chapter 4), and as such, cannot provide conclusive evidence on causality.

CRP concentrations are also lowered by drugs such as statins, which have known efficacy in coronary event reduction. This has also been taken by some as evidence of an aetiological role for CRP in CHD. However, it is possible that the CRP reduction is consequent on, rather than independent of the primary reduction in cholesterol. Therefore, the precise role (if any) of CRP in the process of atherosclerosis is still uncertain.

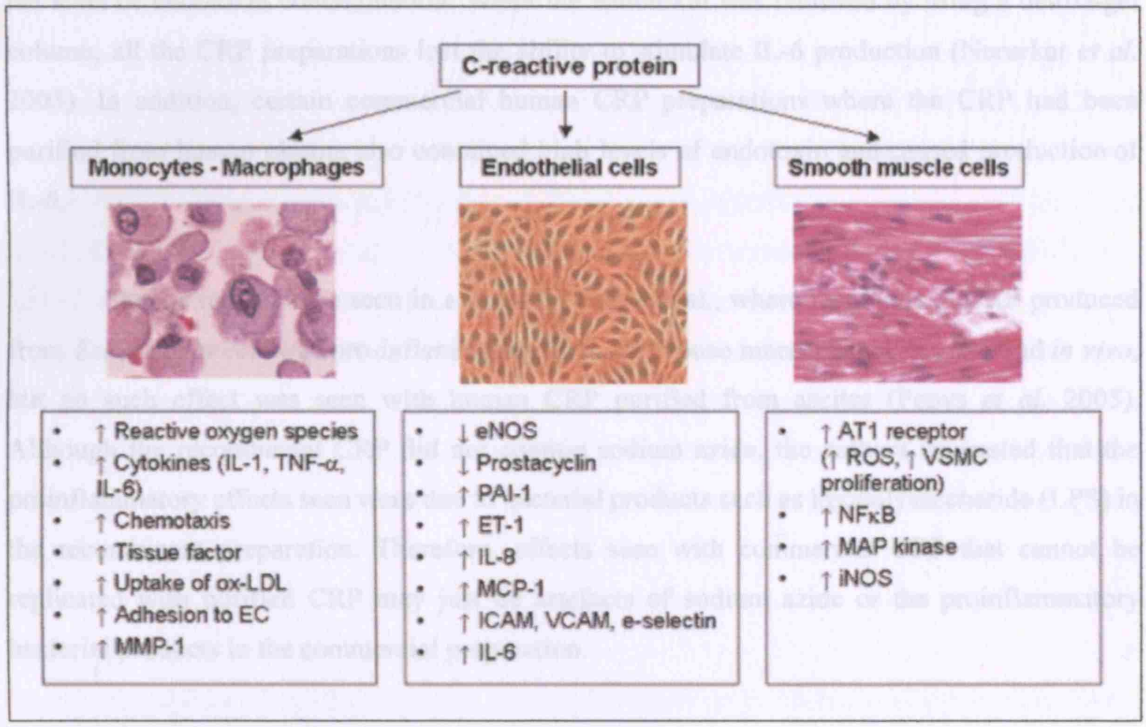
1.3.6 *In vitro* studies investigating a potentially proatherogenic role of CRP

Several *in vitro* studies have been conducted to try and determine whether CRP could have proatherogenic actions. These have shown that CRP might elicit several effects that induce a proatherogenic and pro-inflammatory phenotype in vascular cells and tissues, such as the expression of adhesion molecules, the upregulation of transcription factors such as NF-κB, and the reduction of nitric oxide production (Pasceri *et al.* 2000; Hattori *et al.* 2003; Verma *et al.* 2002) (see Figure 1.8).

ICAM-1 and VCAM-1 are adhesion molecules that play a key role in facilitating the leukocyte-endothelial interaction, an early step in atherogenesis. CRP induced expression of these adhesion molecules in endothelial cells, suggesting it may convert these cells from a quiescent to activated phenotype capable of capturing circulating inflammatory cells (Pasceri *et al.* 2000). In addition, it has been suggested that CRP functions to upregulate the transcription factor NF-κB (Hattori *et al.* 2003), which has been implicated as a key mediator of atherosclerosis. A recent study has suggested that VCAM-1 induction by CRP in endothelial cells may be through NF-κB activation (Kawanami *et al.* 2006). Microarray studies in endothelial cells showed that CRP increased expression of genes such as IL-18 and monocyte chemoattractant protein-1 (MCP1) that are involved in the adhesion of monocytes to vascular endothelium during atherosclerosis, and plasminogen activator inhibitor-1 (PAI-1) that regulates fibrinolysis in atherothrombosis by inhibiting tissue plasminogen activator (tPA), suggesting it is proatherogenic (McCarthy *et al.* 2004; Devaraj *et al.* 2003).

Another potential proatherogenic function of CRP is its ability to reduce the production of nitric oxide from the endothelium, by inhibiting endothelial nitric oxide synthase protein expression, which might block NO-dependent processes, such as angiogenesis and vasodilation that might be atheroprotective (Venugopal *et al.* 2002; Verma *et al.* 2002).

Figure 1.8. Potential proatherogenic effects of CRP on vascular cells.



Complexed CRP has also been thought by some to elicit responses from phagocytic cells through binding to the FcγRI and FcγRIIa receptors (Bharadwaj *et al.* 1999; Marnell *et al.* 1995), however, recent studies using recombinant and highly purified human CRP do not confirm these interactions with human cells (Hundt *et al.* 2001).

Recently, however, the validity of all such experimental studies on vascular cells and tissues has been called into question. Several recent investigations have shown that many of the pro-inflammatory and proatherogenic actions of CRP were in fact mediated by the preservative sodium azide in commercial preparations of CRP that were used in many of the experiments (Pepys *et al.* 2005). A recent study by Van Den Berg et al. showed that commercial CRP induced vasorelaxation in vascular smooth muscle cells, which was then abolished when sodium azide was removed from the preparation by dialysis, indicating that the vasorelaxation was due to sodium azide and not to CRP itself (Van Den Berg *et al.* 2004).

Moreover, since commercial preparations of CRP are synthesised by recombinant technology in bacterial cells, the apparent effects of CRP on vascular cells and tissues might actually be the result of contamination with bacterial endotoxins that are known to exert profound deleterious effects on the vasculature (Bhagat *et al.* 1996). A study by Nerurkar et al.

found that batches of *E. coli* derived recombinant CRP that were able to stimulate IL-6 in endothelial cells, contained endotoxin. The CRP-induced secretion of IL-6 was correlated with the level of endotoxin contamination. When the endotoxin was removed by using a detoxi-gel column, all the CRP preparations lost the ability to stimulate IL-6 production (Nerurkar *et al.* 2005). In addition, certain commercial human CRP preparations where the CRP had been purified from human plasma also contained high levels of endotoxin and caused production of IL-6.

Similar results were seen in a study by Pepys et al., where recombinant CRP produced from *Escherichia coli* was pro-inflammatory both for mouse macrophages *in vitro* and *in vivo*, but no such effect was seen with human CRP purified from ascites (Pepys *et al.* 2005). Although the recombinant CRP did not contain sodium azide, the authors suggested that the proinflammatory effects seen were due to bacterial products such as lipopolysaccharide (LPS) in the recombinant preparation. Therefore, effects seen with commercial CRP that cannot be replicated with purified CRP may just be artefacts of sodium azide or the proinflammatory bacterial products in the commercial preparation.

Nevertheless, some authors claim that CRP retains proatherogenic actions even after using processes such as dialysis or gel filtration to remove contaminants such as sodium azide and bacterial endotoxin (Bisoendial *et al.* 2005). However, this may be because LPS, in particular, binds strongly to proteins and surfaces, and is difficult to remove from biological systems, or that there are other bacterial products in the CRP preparation that are not removed by these purification processes (Pepys *et al.* 2005). Further work is required using purified CRP that does not contain bacterial contaminants to determine if CRP has a proatherogenic role *in vitro*.

An alternative approach that may shed better light on the functions of CRP would be the development of a potent, selective inhibitor. In the past, the development of such an inhibitor has been hampered by a limited understanding of the cellular target or 'receptor' for CRP. Recently, a novel inhibitor has been designed that takes advantage of the fact that CRP potently binds phosphocholine and comprises two phosphocholine molecules linked by their tails (Pepys *et al.* 2006). This inhibitor cross-links two molecules of CRP via their functional phosphocholine-binding surface, thus rendering them inert and facilitating their clearance from the circulation. This CRP inhibitor was shown to abolish the increase in infarct size produced by injection of human CRP in rats undergoing acute MI from coronary ligation, suggesting CRP may promote myocardial damage after MI. However, this inhibitor is currently not available for use in human studies and therefore the potentially proatherogenic role of CRP is still unclear.

1.3.7 Determinants of basal CRP concentration

Although CRP is generally regarded as a marker of infection and low-grade inflammation, which has been used to explain its association with coronary disease, there are hormonal, metabolic and other factors that also affect its concentration. Several factors that are associated with increased risk of atherosclerosis are also associated with differences in CRP concentration. These include age (Ford *et al.* 2004), gender (Wener *et al.* 2000), smoking (Ford *et al.* 2004), BMI (Folsom *et al.* 2001), socio-economic status (Ricardo *et al.* 2007), and birthweight (Sattar *et al.* 2004). These associations could contribute to potential confounding in the association between CRP and CHD (see section on Mendelian randomisation and Chapter 10). HRT use has also been found to be associated with elevations of CRP concentration (Barinas-Mitchell *et al.* 2001). A few studies have also been recently conducted examining whether there are ethnic and ancestral differences in CRP concentration (LaMonte *et al.* 2002; Lakoski *et al.* 2005).

Age, gender and smoking

Several studies, such as the National Health and Nutrition Examination Study (NHANES), the Monitoring Trends and Determinants in Cardiovascular Disease (MONICA) study, VERA study and KORA study have shown a trend of higher CRP concentrations with increasing age (Ford *et al.* 2004; Imhof *et al.* 2003). The NHANES cohort was also examined for differences in CRP concentration when divided according to gender. Ford et al. found that the proportion of women not on HRT who had CRP concentrations greater than 3mg/L was significantly higher than the proportion of men in the same category (21.9% of women compared to 16.4% of men, $p<0.001$). This study also examined the effects of smoking and found that smokers had higher CRP levels compared to non-smokers ($p<0.001$) (Ford *et al.* 2003a). This effect was replicated in the National Heart, Lung and Blood Institute (NHLBI) Family Heart Study (Folsom *et al.* 2001).

Insulin resistance

Other studies have also found similar associations, and have adjusted for these factors when examining the relationship between components of the metabolic syndrome and CRP concentration. The NHLBI Family Heart Study (Folsom *et al.* 2001), NHANES III study (Ford 1999) and the Insulin Resistance Atherosclerosis Study (IRAS) (Festa *et al.* 2001) have both found association between BMI and CRP concentrations ($p<0.0001$) and this is replicated in numerous other studies. In addition, the NHLBI Family Heart Study has investigated the relationship between CRP concentrations and insulin, and diabetes. They found increasing insulin levels are associated with increased CRP levels ($p<0.0001$) in both men and women. With the relationship between diabetes and CRP concentration, there was strong association in

women ($p<0.0001$) and a weaker association in men ($p<0.04$) (Folsom *et al.* 2001). Similarly, a recent study on CRP and the metabolic syndrome found strong associations between BMI and CRP ($p<0.0001$) and insulin resistance and CRP ($p<0.0001$) (Timpson *et al.* 2005).

Socio-economic status

Several studies, such as Whitehall II, the Finnish platelet aggregation and inflammation study, the ATTICA study, the British Women's Heart and Health study, the Young Finns study, the fourth NHANES study and the ARIC study have all shown that the highest CRP concentrations are seen in those from the lowest socio-economic class, and there is an inverse association between the two factors (Owen *et al.* 2003; Jousilahti *et al.* 2003; Panagiotakos *et al.* 2004; Lawlor *et al.* 2005; Kivimaki *et al.* 2005; Alley *et al.* 2006; Ricardo *et al.* 2007). However, where other inflammatory markers have been measured, the same associations have been seen, suggesting the association is not specific to CRP.

Birth weight

Although much data is available on the association between CRP and socio-economic status, far fewer studies have been conducted to examine the association between CRP and birth weight. The MIDSPAN family study examined 1663 individuals with CRP measures who had birth weight data available (Sattar *et al.* 2004). After adjusting for factors known to influence CRP concentrations including age, BMI, smoking, socio-economic status, and hormone use in women, the relationship between CRP and birth weight was assessed. Again, an inverse association was seen, with lower birth weights being associated with higher CRP concentration. However, further studies are required to confirm this association, and evaluate its independence.

Hormone replacement therapy

HRT use is associated with serum levels of CRP in healthy menopausal women. The Cardiovascular Health Study found that women taking oestrogen had 59% higher CRP concentrations compared to non-users of HRT (Cushman *et al.* 1999a). The Women's Health Study (WHS) also showed higher CRP concentrations in women taking oestrogen ($p<0.003$) and women taking oestrogen and progesterone ($p<0.03$), even after adjustment for BMI and age (Ridker *et al.* 1999). In the Healthy Women Study, BMI and HRT were both associated with increased levels of CRP ($p<0.001$); CRP levels were 3.01mg/L in women on HRT compared to 1.74mg/L in those not on HRT (Barinas-Mitchell *et al.* 2001). There appears to be a clear effect of HRT on CRP, however, whether this effect is specific to CRP or is a general effect on inflammation is unclear.

Ethnic differences

Risk factors such as smoking, BMI and adiposity also show characteristic patterns of clustering across different ethnic backgrounds and could account in part for differences in CRP

concentration observed between populations of differing ethnicity and ancestry (Harland *et al.* 1997; Hunt *et al.* 2002; Howard *et al.* 2003; Forouhi & Sattar 2006). However, studies comparing CRP concentrations in different populations have found that the differences persist even after adjustment for these factors, that might reflect genetic differences in CRP regulation, or the presence of residual confounding (Chambers *et al.* 2001; Forouhi *et al.* 2001; LaMonte *et al.* 2002; Chandalia *et al.* 2003).

These associations are important to evaluate in more detail since they may be critical when trying to understand the aetiological role of CRP in coronary disease. They suggest that the association seen between CRP and coronary disease in observational studies has substantial potential for confounding compromising efforts to establish the causal relevance of CRP in coronary risk. It may be important to adjust for these factors when examining the relationship between CRP and CHD. However, unless the pathway linking CRP to increasing CHD risk is known, there is also the possibility of over-adjustment. For example, if some of these factors such as insulin resistance, diabetes or increased blood pressure mediated the effect of CRP on coronary disease, it would be inappropriate to adjust for them.

1.3.8 Biology of CRP transcription

CRP concentration can rise from low basal levels to very high levels in infection and inflammation. These differences in concentration may, in theory, reflect increased synthesis or decreased degradation of CRP. However, it is known that clearance of CRP is constant with a half-life of 19 hours, irrespective of basal state or infection (Pepys 1982). Therefore, active regulation of CRP is mostly at the level of synthesis, and since CRP is not stored, this is likely to be mediated by altering the rate of transcription of the CRP gene, and perhaps also by modification of mRNA stability.

The CRP transcript is characterised by the presence of a 1.2 kb, 3' untranslated region (UTR), which is thought to mediate its rapid degradation following reduction of inflammation (Volanakis 2001). Studies in transgenic mice have also shown that increasing the length of either the 5`- or 3`-flanking DNA decreases the basal level of expression, indicating the presence of regulatory regions in these sequences (Murphy *et al.* 1995). Transcriptional regulation of CRP has been studied extensively both *in vitro* and *in vivo*. IL-6 appears to be the principal inducer of CRP gene transcription, while IL-1 and other factors such as complement activation products, act synergistically with IL-6 to enhance its effect (Toniatti *et al.* 1990b; Toniatti *et al.* 1990a; Ganapathi *et al.* 1991; Szalai *et al.* 2000a).

1.3.9 Transcription factors involved in CRP expression

Several transcription factors have been found to participate in the regulation of CRP gene transcription. Early studies carried out by Majello et al. demonstrate that the CRP promoter contains two adjacent binding sites (β and α) within 90 bases of the transcriptional start site, that interact with at least two hepatocyte-specific nuclear proteins, H-APF-1 and H-APF-2. Mutations that abolish or reduce binding at these sites reduce the level of CRP gene expression (Majello *et al.* 1990). *In vitro* studies carried out in a human hepatoma cell culture system have attempted to identify cytokine responsive elements that result in inducible CRP gene expression. Studies by Li et al. found that there were two regions of these inducible elements between –88bp and –60bp and between –234bp and –200bp, later termed acute phase responsive elements (APREs). They also found two constitutive enhancer-like regions between –339bp and –292bp and between –904bp and –835bp, and a negative regulatory region between –178bp and –88bp (Li & Goldman 1996). This confirmed work carried out by Arcone et al., who reported an inducible element between –121bp and –50bp (Arcone *et al.* 1988).

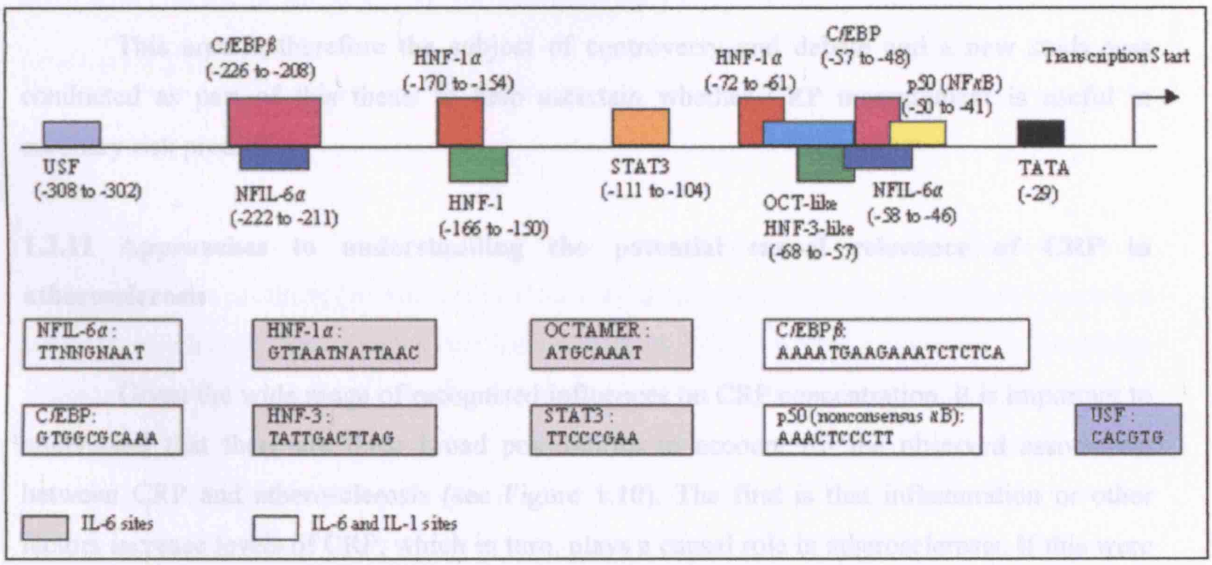
It is known that the APREs contain the necessary information for liver specific and IL-6 inducible expression in human hepatoma Hep3B cells (Toniatti *et al.* 1990a). Further studies by the same group found that both APREs contain a low-affinity binding site for the liver specific transcription factor, hepatocyte nuclear factor-1 (HNF-1), involved in developmentally regulated gene expression in the liver. The two sites are separated by approximately 80bp and experimental mutations in either site abolish inducible expression (see Figure 1.9). Moreover, an expression vector encoding HNF-1 was capable of trans-activating transcription from the wild-type CRP promoter but not from mutants that have lost the ability to bind HNF-1. This suggests that two HNF-1 molecules bound simultaneously to sites distant from each other can act synergistically to activate gene expression (Toniatti *et al.* 1990b).

Various members of the CCAAT-enhancer binding protein (C/EBP) family bind to APREs, in particular, IL-6REs. Initial attempts to identify which members are involved in IL-6 signal transduction led to the isolation of the transcription factor C/EBPβ, which is activated post-translationally by IL-6 (Poli *et al.* 1990; Cao *et al.* 1991). Another member of the C/EBP family was later identified and termed C/EBPδ. This transcription factor differs from C/EBPβ in that it is induced at the transcriptional level (Ramji *et al.* 1993). C/EBPβ activates expression of C/EBPα and peroxisome proliferator-activated receptor-γ (PPARγ), which then transcriptionally activates genes that produce an adipocytic phenotype. Findings by Tang et al. demonstrate that C/EBPβ is required for mitotic clonal expansion in adipocyte differentiation (Tang *et al.* 2003). C/EBPβ and δ have been found to participate in CRP expression and bind to two sites on the proximal part of the CRP promoter (Agrawal *et al.* 2001; Cha-Molstad *et al.* 2000). Nuclear

factor interleukin-6α (NFIL-6α) is another member of the C/EBP family, and is dramatically induced by inflammatory cytokines such as IL-1 and IL-6 (Isshiki *et al.* 1991). Data from Li et al. shows that both IL-6REs contain an NFIL-6α binding site and that the downstream IL-6RE also contains a positive HNF-1α binding site and a negative HNF-3/Octamer-like factor binding site (Li & Goldman 1996) (see Figure 1.9).

Another family, the signal transducers and activators of transcription (STAT) family have also been shown to be important mediators of the effects of many cytokines including IL-6. A study by Zhang et al. has found that STAT3 binds to a site on the CRP gene at –108bp, and that it participates in the transcriptional activation of CRP in response to IL-6 (Zhang *et al.* 1996). Rel p50, a member of the Rel family, has also been studied (Cha-Molstad *et al.* 2000) and binds to a nonconsensus κB site centered at –46bp, overlapping the proximal C/EBP site (see Figure 1.9). The findings from this group suggest that IL-1β induces nuclear translocation of p50-containing dimers and that p50 interacts with C/EBPβ activated by both IL-6 and IL-1β to induce CRP expression. The overlapping C/EBP binding and nonconsensus κB sites have been shown to be necessary for maximal CRP expression in response to cytokines (Cha-Molstad *et al.* 2000; Agrawal *et al.* 2001). Recent data has indicated that c-Rel (required for C/EBPβ binding to the promoter) regulates CRP gene expression by a novel mechanism, without the requirement of binding to DNA/κB site, and binds directly to C/EBPβ to facilitate the binding of C/EBPβ to the promoter (Agrawal *et al.* 2003).

Figure 1.9. CRP promoter region showing transcription factor binding sites, the sequences of the transcription factors and whether they are induced by IL-6, IL-1 or both.



Recent molecular studies indicate that CRP transcription is influenced not only by inflammatory cytokines such as IL-6 and IL-1, but also by transcriptional regulators involved in the regulation of lipid and metabolic genes such as USF1 (Corre & Galibert 2005), as well as by endoplasmic reticulum stress pathways that include transcription factors such as CREBH (Zhang *et al.* 2006a). Understanding the mechanisms by which these transcriptional mechanisms are activated to influence CRP gene transcription could help to explain the variation in CRP concentrations seen in healthy individuals in the absence of overt inflammatory stimuli and perhaps explain the association of CRP with insulin and adiposity.

1.3.10 Potential clinical use of CRP in the prediction of cardiovascular events

Aside from understanding a potential causal effect of CRP in atherosclerosis, there has been interest in the use of CRP measurement to evaluate risk of future cardiovascular events. Studies by Ridker et al., Koenig et al. and several others arrived at the conclusion that CRP adds prognostic information to risk scores such as the Framingham risk score, that utilize established cardiovascular risk factors such as smoking, blood pressure, diabetes and cholesterol (Ridker *et al.* 2002; Ridker 2003b; Koenig *et al.* 1999; Koenig 2005). Other studies, such as those by Danesh et al. and Wilson et al. suggest that once established risk factors are taken into account, the incremental predictive utility of CRP may be limited and that it may not add information to the Framingham risk score (Danesh *et al.* 2004; Wilson *et al.* 2005). Another paper by Ridker suggests that CRP is a better predictor of cardiovascular events than LDL-cholesterol, and argues in favour of the use of CRP screening in primary prevention to identify high-risk patients not identified from cholesterol screening and treat these individuals with statins to lower risk (Ridker 2003c).

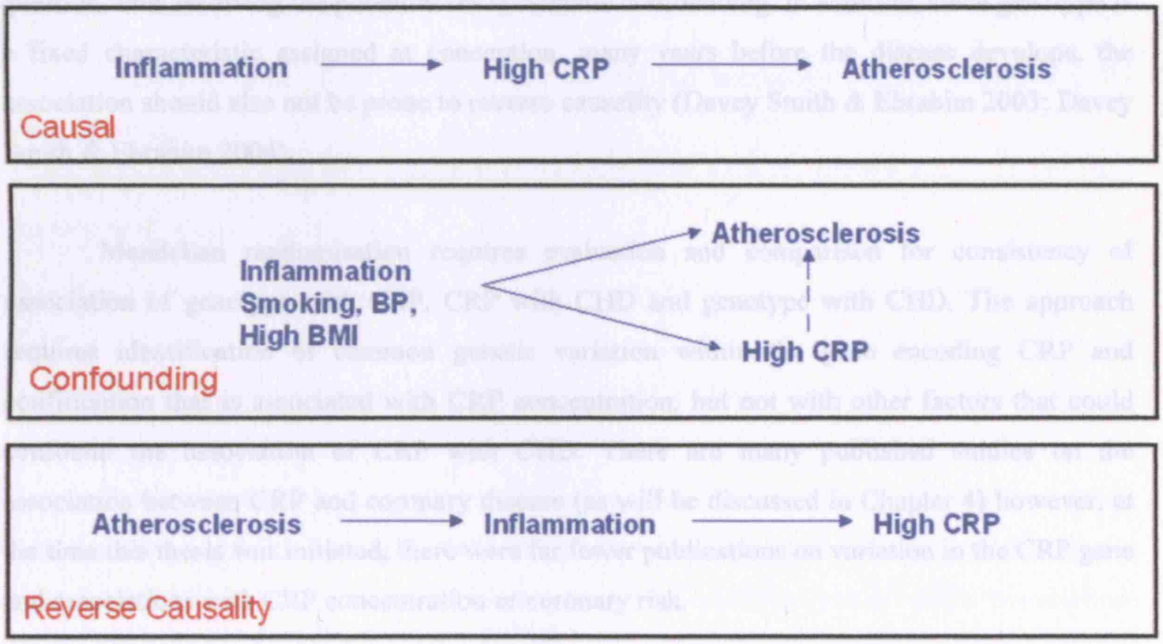
This area is therefore the subject of controversy and debate and a new study was conducted as part of this thesis to help ascertain whether CRP measurement is useful in coronary risk prediction.

1.3.11 Approaches to understanding the potential causal relevance of CRP in atherosclerosis

Given the wide range of recognised influences on CRP concentration, it is important to understand that there are three broad possibilities to account for the observed association between CRP and atherosclerosis (see Figure 1.10). The first is that inflammation or other factors increase levels of CRP, which in turn, plays a causal role in atherosclerosis. If this were the case, CRP would be a therapeutic target, and lowering CRP would be expected to be associated with a reduction in coronary events. Alternatively, CRP may simply be a marker of

other inflammatory processes, or of established risk factors such as blood pressure, cholesterol or smoking that are causally linked to CHD. In this case, the association of CRP with CHD risk would be the result of confounding and lowering CRP therapeutically would not lead to reductions in CHD events. The third possibility is that the elevations of CRP reflect inflammatory processes occurring within the atherosclerotic plaque rather than being a cause of them (so-called reverse causality). Although prospective studies of initially healthy individuals should be less prone to the effects of reverse causality than case-control studies, because atherosclerosis has a long subclinical phase, with lesions being established even by the second decade (McGill *et al.* 2000), reverse causality remains a potential explanation of the observed associations.

Figure 1.10. Links between CRP concentration and atherosclerosis.



Similar uncertainties arising from associations of blood pressure and cholesterol with coronary disease in observational studies were eventually resolved by randomised controlled trials, in which the levels of these risk factors were modified by drug treatment. Such a design balances known and unknown confounders between the treatment and placebo groups. Any effect of treatment on disease outcome should therefore be through the risk factor that is modified, providing evidence for a causal role for the risk factor. Clinical trials of the selective CRP lowering agent are likely to be some way off (Pepys *et al.* 2006). In the meantime, an emerging and complementary alternative is to use a genetic approach termed "Mendelian randomisation", which will be discussed in more detail in the next section.

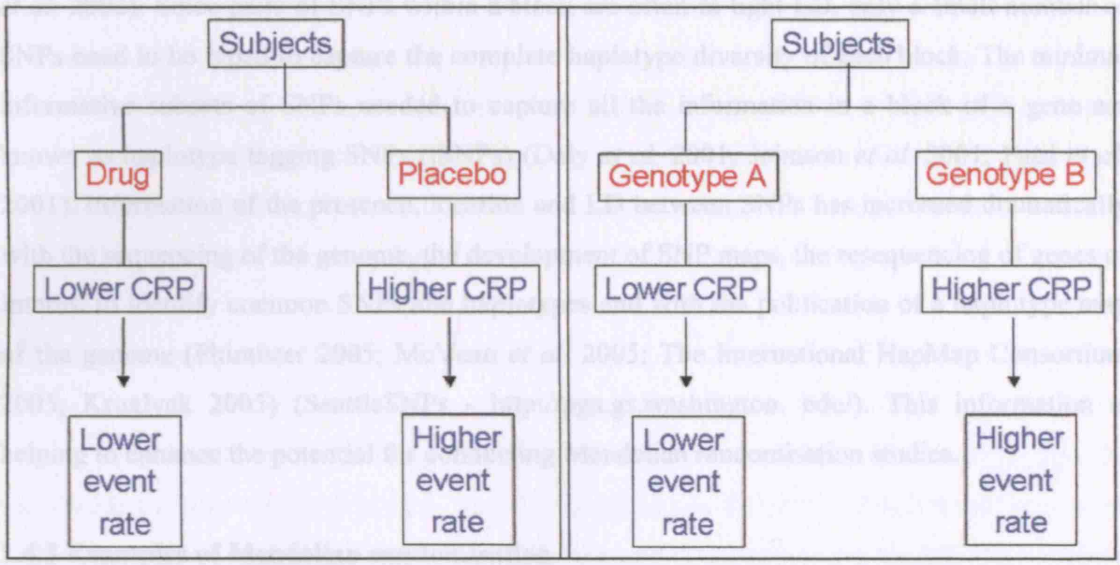
1.4 Mendelian randomisation

1.4.1 Principles of Mendelian randomisation

Although bias and confounding can be overcome in RCTs, for many putative factors such as IL-6, fibrinogen or CRP, drugs that selectively alter plasma levels and can be used in humans are not currently available so an intervention trial is not yet possible. In this situation, genetic studies may offer an alternative randomised design that helps to minimise confounding. Analogous to RCTs where the allocation to the treatment or control group is random, alleles in genes that affect the levels of a factor of interest (such as CRP) are also assigned randomly at conception, according to Mendel's law of independent assortment of alleles (Mendel 1865; Davey Smith & Egger 1998) (see Figure 1.11). This means that potential confounders, whether known or unknown, should be distributed evenly among carriers and non-carriers of the allele in question, thus resolving the potential for systematic confounding. In addition, since genotype is a fixed characteristic assigned at conception, many years before the disease develops, the association should also not be prone to reverse causality (Davey Smith & Ebrahim 2003; Davey Smith & Ebrahim 2004).

Mendelian randomisation requires evaluation and comparison for consistency of association of genotype with CRP, CRP with CHD and genotype with CHD. The approach requires identification of common genetic variation within the gene encoding CRP and confirmation that is associated with CRP concentration, but not with other factors that could confound the association of CRP with CHD. There are many published studies on the association between CRP and coronary disease (as will be discussed in Chapter 4) however, at the time this thesis was initiated, there were far fewer publications on variation in the CRP gene and associations with CRP concentration or coronary risk.

Figure 1.11. Mendelian principles to determine causality and similarities with a randomised controlled trial.



Left: Expected outcome from hypothesised randomised controlled trial of selective CRP-lowering drug.
Right: Expected outcome from Mendelian randomisation analysis if CRP were causal in the development of coronary events.

1.4.2 Common genetic variation in the human genome

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation in the human genome (Collins *et al.* 1997). They are stable sequence variations, in which usually two alternate nucleotide bases are observed at one position across some populations. Insertion/deletion polymorphisms can also occur, where bases are inserted or deleted into the genome sequence. Polymorphisms can occur within coding regions of a gene and may alter the size or shape of the resulting protein. SNPs can also occur in non-coding regions and may alter the expression or stability of a protein. Polymorphisms therefore have the ability to alter the function or activity of a gene. Even though most SNPs have no observable phenotype, they may have now been linked to variation in the levels of the trait of the factor they encode and, increasingly now with risk of complex diseases such as T2DM (Pro12Ala in the PPAR-γ gene) and age-related macular degeneration (Tyr402His in the CFH gene) (Douglas *et al.* 2001; Maller *et al.* 2006).

Sets of nearby SNPs on the same chromosome or in the same gene are often inherited together in blocks. Such SNPs are said to be in linkage disequilibrium (LD), with a particular

pattern of alleles on any one chromosome being termed the haplotype. Each block normally contains only a few common haplotypes (Stephens *et al.* 2001b; Stephens *et al.* 2001a; Gabriel *et al.* 2002). Since pairs of SNPs within a block are often in tight LD, only a small number of SNPs need to be typed to capture the complete haplotype diversity of each block. The minimal informative subsets of SNPs needed to capture all the information in a block of a gene are known as haplotype tagging SNPs (tSNPs) (Daly *et al.* 2001; Johnson *et al.* 2001; Patil *et al.* 2001). Information of the presence, location and LD between SNPs has increased dramatically with the sequencing of the genome, the development of SNP maps, the resequencing of genes of interest to identify common SNPs and haplotypes and with the publication of a haplotype map of the genome (Phimister 2005; McVean *et al.* 2005; The International HapMap Consortium 2005; Kruglyak 2005) (SeattleSNPs - http://pga.gs.washington. edu/). This information is helping to enhance the potential for conducting Mendelian randomisation studies.

1.4.3 Examples of Mendelian randomisation

The use of Mendelian randomisation to overcome confounding and make inferences about the causes of a disease is a novel but increasingly established concept. Several studies have already been carried out examining the effects of circulating factors that have been implicated in CHD. Mutations in the APOB gene, in particular the Arg3500Gln mutation gives rise to higher levels of total cholesterol in heterozygotes (2.6mmol/L higher), without an effect on triglycerides, fibrinogen or BMI. In genetic studies, heterozygotes for this variant had an OR of 7.0 for CHD compared with the general population, consistent with the now established causal link between cholesterol and CHD risk (Tybjarg-Hansen *et al.* 1998).

The association between homocysteine and CHD has also been examined using Mendelian randomisation. This relationship may be confounded in observational studies by factors such as smoking and blood pressure, with which homocysteine is also associated. However, polymorphisms exist in the methyltetrahydrofolate reductase (MTHFR) gene that are associated with differences in serum homocysteine levels. The 677TT genotype is associated with homocysteine levels that are higher by around 1.7μmol/L compared to individuals homozygous for the C-allele. In a meta-analysis of published studies, TT subjects had an OR for stroke of 1.26 (95%CI: 1.14-1.40) compared to CC subjects, providing evidence of a causal link between homocysteine and stroke (Casas *et al.* 2005). The use of Mendelian randomisation has also been used to investigate the link between homocysteine and CHD, but the interpretation of the results have been conflicting (Wald *et al.* 2002; Klerk *et al.* 2002; Lewis *et al.* 2005). Randomised controlled trials of homocysteine lowering with folic acid are underway and will help to confirm or refute the findings of Mendelian randomisation studies.

Mendelian randomisation has also been used to show that fibrinogen is unlikely to be a causal factor in CHD, even though case-control studies have shown a 0.12g/L increase in fibrinogen is associated with a relative risk of 1.20 for CHD. A polymorphism in the promoter region of the β-fibrinogen gene was shown to alter plasma fibrinogen levels by 0.12g/L. Therefore, if fibrinogen is a causal factor in CHD, individuals with the polymorphism should have a relative risk of 1.2 for CHD. However, the OR for CHD associated with the variant in a meta-analysis was 1.00 (95%CI: 0.95-1.04), suggesting that the relationship between fibrinogen and CHD is confounded (Davey Smith *et al.* 2005a; Keavney *et al.* 2006).

1.4.4 Genetic studies of CRP

Identification of common variants in the CRP gene that affect its level is a pre-requisite for Mendelian randomisation studies. The first insights into the potential genetic influences on CRP came from heritability studies. Familial correlation analyses in the NHLBI Family Heart Study gave a coefficient value of 0.21 between siblings and 0.18 between parent and child (Pankow *et al.* 2001). However, in family studies, it is difficult to distinguish how much correlation is due to shared familial environment and how much is due to shared genes. Twin studies are important as they allow correlation to be studied in monozygotic and dizygotic twins that share 100% and 50% of their genes respectively, allowing the correlation due to environment and genes to be explored.

To date, only four twin studies have been carried out investigating the role of genetics of CRP in a low-grade inflammatory state. A comparison of monozygotic and dizygotic twins found a significantly higher correlation of CRP among monozygotic than among dizygotic twins, with an estimated heritability of baseline values being 52%, thus showing a significant genetic contribution to the control of baseline CRP concentrations (MacGregor *et al.* 2003). A second study carried out in healthy male and female monozygotic twins produced a within-pair correlation coefficient of CRP of 0.40, therefore also suggesting a significant degree of heritability and hence an important genetic component (Retterstol *et al.* 2003). Another study carried out in female monozygotic twins, found that 54% of variation in circulating CRP was due to genetic influences (Greenfield *et al.* 2004). The most recent study, however, resulted in a heritability coefficient of 0.40, which was attenuated to 0.22 after adjustment for covariates such as age, gender, BMI, alcohol intake and smoking status, suggesting the genetic contribution may be smaller than previously thought (Worns *et al.* 2006). However, it is important to note that the analyses have differed in these studies; therefore it is not possible to directly compare the heritability values.

1.4.5 Polymorphisms in the CRP gene

Initial sequencing of around 3050bp containing the CRP gene identified a multiallelic GT repeat polymorphism within the intron, which can be repeated between 15-24 times (Goldman *et al.* 1987; Weber *et al.* 1990). Later studies investigated whether this dinucleotide repeat polymorphism was associated with variation in baseline CRP and found significant association of length variation with CRP concentration (Szalai *et al.* 2002).

A synonymous coding region polymorphism (+1059G/C) in exon 2 was also reported (Cao & Hegele 2000) and it was later shown to be associated with differences in CRP concentration (C-allele carriers having lower concentrations) (Russell *et al.* 2003). No association was identified of this polymorphism with arterial thrombosis (Zee & Ridker 2002). A +2302G/A variant downstream of the 3'UTR was also associated with reduced CRP concentrations (Russell *et al.* 2003). This difference was more marked in individuals with both the +2302AA and +1059CC genotypes. However, at the time this thesis was started, no association had been identified between the +2302G/A polymorphism and risk of vascular disease.

In an association study of Type II diabetes conducted in Pima Indians, six polymorphisms in and around the CRP gene were identified (Wolford *et al.* 2003). These included two variants in the promoter region (-717A/G and –286C/T/A), one in the intron (+194A/T) and three in the 3'UTR (+1444C/T, +2007T/C and +2302G/A). Out of these SNPs, the –717A/G and +1444C/T were assessed for association with diabetes and the A-allele of the –717A/G polymorphism was found to be associated with an increased prevalence of diabetes in Pima Indians.

Studies carried out by Brull et al. using a single strand confirmation polymorphism (SSCP) screen also identified the –717A/G promoter variant and the +1444C/T polymorphism in the 3'UTR, which were confirmed by sequencing (Brull *et al.* 2003). The +1444C/T polymorphism does not appear to disrupt any known consensus sequences associated with altered mRNA stability (Russell *et al.* 2003). The +1444C/T variant was found to be associated with CRP both basally and after an acute inflammatory response, whereas no significant effect on CRP levels was seen with the –717A/G promoter variant (Brull *et al.* 2003).

Taken together, these studies indicate that CRP concentration in part, is regulated by common variants in the CRP gene. With the availability of public domain resources such as NCBI dbSNP and HapMap, and resequencing databases such as SeattleSNPs that contain information on CRP variants, a consensus map of CRP variation was generated (see Chapter 5).

A systematic overview of several of these SNPs and their effects on CRP concentration was conducted as part of this thesis as more studies of the associations were published to obtain the precise estimates required for this arm of a Mendelian randomisation analysis (see Chapter 6).

1.4.6 Limitations of Mendelian randomisation

Although the principles of Mendelian randomisation could, in theory, be used to determine if an exposure like CRP is causally involved in CHD, the approach may have limitations. One of the earlier limitations of the Mendelian randomisation approach was a lack of suitable polymorphisms for examining the exposure of interest, even if a locus was involved in a disease-related process (Davey Smith & Ebrahim 2004). However, recent public domain resources such as NCBI dbSNP, ENSEMBL, HapMap, and SeattleSNPs aim to overcome this limitation by resequencing of genes and production of high and low density SNP maps containing both common and rare variants, many of which have since been found to be associated with intermediate phenotypes.

Although using the Mendelian randomisation approach is a means to reduce confounding, it may not be removed completely. If alleles at the locus being studied are in LD with variants at another adjacent gene, the effect of the polymorphism under investigation may be confounded by the influence of the other polymorphism. Another problem that could arise with investigating a single polymorphism is that it may have pleiotropic effects, by affecting more than one intermediate phenotype of interest. This would almost certainly require that the gene being studied is alternatively spliced. However, this may be corrected for by measuring the other intermediate phenotypes and adjusting for them. Confounding can also occur if the polymorphism studied is associated with ethnicity and ethnicity is also associated with an increased risk of disease. One way to reduce this problem is to restrict studies to a single homogeneous population.

If the associations between genotype and a potential intermediate phenotype, or between genotype and disease outcome, are not reliably estimated, then interpretation of the consistency of the associations will be inaccurate. One of the disadvantages of gene-disease association studies is that the results can be inconsistent and difficult to replicate, often because of inadequate sample sizes and the propensity for false positive or negative studies. Meta-analyses of such studies may enhance the reliability of the genetic effect size but if such analyses are restricted to published studies, they may be prone to the effects of publication bias.

Meta-analyses can be used to combine genetic association studies, where again, lack of power is a problem. To obtain a typical odds ratio of 1.3, sample sizes of 2000-10000 cases and

controls are required to obtain 80% power (Zondervan & Cardon 2004), therefore, many studies may fail to reach statistical significance. By combining genetic studies, it may be possible to provide stronger evidence of a small effect (Munafo & Flint 2004). An example of this can be seen with the association between APOE4 and Alzheimer's disease, where the allele frequency is 0.15 and the genotype relative risk between E4E4 and E3E3 homozygotes is 11.57, found from a meta-analysis of 42 case-control studies (Rubinsztein & Easton 1999).

In meta-analyses where small studies have been combined, it is important to test for publication bias. An example of where the results of meta-analyses have been overturned due to publication bias is the angiotensin converting enzyme (ACE) insertion/deletion (I/D) polymorphism, where meta-analyses of small studies have been contradicted by large RCTs (Borzak & Ridker 1995; Cappelleri *et al.* 1996; Ioannidis *et al.* 1998). The ACE I/D polymorphism is associated with plasma ACE activity and small studies have found it to be associated with an increase in risk of MI. However, large studies have seen little or no association between this polymorphism and increased risk of MI. Later assessment of meta-analyses of the small studies has found evidence of publication bias (Samani *et al.* 1996; Staessen *et al.* 1997). This suggests selective non-publication of studies with negative results. Large RCTs and recent meta-analyses of large studies suggests that the association between the ACE I/D polymorphism and increased risk of MI may not be real (Agerholm-Larsen *et al.* 1997; Agerholm-Larsen *et al.* 2000).

Confounding and bias can also be present in studies where the population is stratified according to disease risk and allele frequencies at the marker locus, giving rise to false-positive results (Tobin *et al.* 2004). Nevertheless, the pooling of data in such meta-analyses is likely to be central to the conduct of such studies (Ioannidis *et al.* 2006).

Genetic studies and meta-analyses of genotype and intermediate phenotype associations are likely to be much less affected by inconsistency arising from small sample sizes. This is because intermediate phenotypes are continuous traits, so the power of any study is increased compared to studies where the outcome is categorical (e.g. disease event). They are also less prone to misclassification. Also, these intermediate phenotypes are usually the product of the gene itself, and are a more immediate consequence of genetic variation and therefore influenced by a smaller range of modifying genes and exposures than the disease, so the signal-noise ratio in genotype-intermediate phenotype studies is likely to be much greater. There are now many examples of established genotype-intermediate phenotype associations for example for triglycerides, HDL- cholesterol, fibrinogen and homocysteine.

Hypothesis

2.1 Hypothesis and central question

Given the observed associations between CRP concentrations and coronary heart disease and between CRP polymorphisms and CRP concentrations from published studies, I developed the following hypotheses:

1. CRP concentrations are predictive in the detection of coronary heart disease events in healthy individuals.
2. CRP polymorphisms are robustly associated with differences in CRP concentration, both basally and in the presence of infection or inflammation.
3. Differences in CRP concentrations in populations of different ancestry could be explained by differences in frequency of CRP polymorphisms, or their effect on CRP concentrations in different populations.
4. The association between CRP genotype and clinical coronary events should not be subject to confounding by classical or novel risk factors for atherosclerosis.
5. Comparison of risk estimates from genetic and non-genetic studies of CRP and clinical events will provide insight into the nature of the CRP-coronary event association.

2.2 Aims

The major aim of this work was to identify common genetic variants in the CRP gene that influence CRP concentration and to assess whether the CRP-IHD association is causal or associative using a Mendelian randomisation approach. Additional aims were to investigate whether CRP concentrations provided useful information in determining risk of coronary disease.

Specific aims:

1. Evaluate the performance of CRP in the detection of later coronary heart disease events in comparison with other risk factors and biomarkers based on hypothesis 1 (see Chapter 4).
2. Establish a CRP gene map to locate all the polymorphisms in the gene by collating and comparing data from public domain databases of sequence variation based on hypothesis 2 (see Chapter 5).
3. Examine linkage disequilibrium within the CRP gene to facilitate the generation of tagging SNPs for haplotype-CRP association studies based on hypothesis 2 (see Chapter 5).
4. Investigate the precise effect of CRP polymorphisms on CRP concentrations both from published data and new studies and assess whether the association is modified by a variety of exposures relevant to coronary risk based on hypotheses 2 and 4 (see Chapters 6 and 7).

5. Investigate whether genotype influenced the CRP fluctuations during an inflammatory episode based on hypothesis 2 (see Chapter 8).
6. Assess whether ethnicity and ancestry influence population variation of CRP concentration, and whether this might be explained by differences in presence or frequency of CRP genotypes and/or their effect on CRP concentration based on hypothesis 3 (see Chapter 9).
7. Investigate the effect of the +1444C/T genotype on future risk of vascular disease using a Mendelian randomisation approach. Risk estimates from the genetic approach will be compared with risk estimates from observational studies to determine if CRP plays a causal role in CHD or if any association seen between CRP and CHD is subject to confounding based on hypotheses 4 and 5 (see Chapter 10).

Methods

3.1 Public domain resources

3.1.1 Mining public domain databases for SNPs in the CRP gene

Although a number of public domain resources now hold information on polymorphic variation in the human genome, the means by which this information is obtained varies, with the result that each database contains partially overlapping information. Several public domain databases were therefore searched with the aim of constructing a consensus map of polymorphisms in the CRP gene and its flanking regions.

The NCBI Single Nucleotide Polymorphism database (dbSNP) (http://www.ncbi.nlm.nih.gov/SNP/) is a collaboration between the National Human Genome Research Institute and the National Center for Biotechnology Information, and serves as a public-domain archive for a broad collection of simple genetic polymorphisms. The NCBI RefSeq collection contains non-redundant sequences including genomic DNA, gene transcripts (RNA), and protein products. The complete CRP gene sequence and its pseudogene sequence (accession number AF442818) are available on this site, which can also be accessed via other NCBI databases, such as the Entrez Nucleotide database (http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=17933262). This sequence was submitted by Harraghy et al. in 2001, and is also linked to the NCBI database for SNPs (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=1401), which gives details of confirmed and suspected SNPs within the genomic region of CRP, the average estimated heterozygosity, and the average allele frequency. These values have been estimated by genotyping families from the Centre d'Etude du Polymorphisme Humain (CEPH), consisting of 30 mother-father-child trios that are Utah residents with ancestry from northern and western Europe, and the NIH Polymorphism Discovery Resource (NIHPDR) that consists of DNA from 450 anonymous, unrelated individuals with equal numbers of females and males.

The NHLBI Programs for Genomic Applications (PGA) SeattleSNPs database (http://pga.gs.washington.edu/data/crp/) also used sequences from the NCBI RefSeq database as a template from which to resequenced the gene in ~1Kb overlapping amplicons to produce a high density SNP-map. This database has identified common variable sites in the CRP gene, and established their relative frequencies in two human populations with different evolutionary histories (African and European). The African descent population consists of 24 individuals (12 male/12 female) and is composed of DNA available from the Coriell Cell Repository (http://locus.umdnj.edu/). These individuals were selected from a human variation panel of 50 African Americans. The European descent population consists of 23 individuals (12 male/11 female) and is composed of Centre d'Etude du Polymorphisme Humain (CEPH) parent DNA. These individuals are Utah residents with ancestry from northern and western Europe. The

database provides a visual genotype for all the SNPs for each individual, and uses this data to obtain the allele frequency, the Hardy-Weinberg probability and pairwise linkage disequilibrium (allelic association) between SNPs. This information is used to infer common haplotypes (see Chapter 5).

The HapMap Consortium database (http://www.hapmap.org) is a culmination of information gathered from genotyping over one million SNPs across the human genome resulting in the production of low-density SNP maps of hundreds of genes across the genome. A total of four geographical populations have been genotyped in 270 individuals. These are 30 trios (both parents and one child) of Yoruba people from Ibadan, Nigeria (YRI), 45 unrelated Japanese people from Tokyo (JPT), 45 unrelated Chinese from Beijing (CHB), and 30 CEPH trios (both parents and one child) (CEU).

The MatInspector program (http://www.genomatix.de/products/MatInspector/index.html) was also used to examine if any polymorphisms in the promoter region were located in transcription factor binding sites, and whether any base changes resulted in gain or loss of binding sites. MatInspector is a software tool that utilises a large library of matrix descriptions for transcription factor binding sites to locate matches in DNA sequences. It assigns a quality rating to matches and therefore allows filtering and selection of matches (Quandt *et al.* 1995). The CRP promoter sequence (~1050bp) was entered into the program and a list of transcription factor binding sites was generated. The sites containing polymorphisms were then compared with each allele to see if the sites were affected by the SNP.

3.1.2. Literature-based searches for systematic reviews

Three electronic databases (PubMed Medline, EMBASE and the Cochrane Collaboration library) were searched for literature-based comparisons. The literature search was limited to "human", and "English Language". Abstracts from major cardiovascular related conferences in the last 5 years were identified. Any additional studies in the references of all identified publications were also searched.

3.2 Studies evaluated

The relationship between CRP genotype, CRP concentration and cardiovascular events was evaluated in a number of prospective, cross-sectional and case-control studies (see Tables 3.1 and 3.2). For the Mendelian randomisation analyses, exclusion criteria also included female gender and non-Caucasian subjects to to ensure homogeneity across studies, since gender and ethnicity have been shown to influence CRP concentration (see Chapter 10).

3.2.1 NPHSII cohort

The Second Northwick Park Heart Study (NPHSII) is a large prospective study of 3,012 healthy Caucasian middle-aged (50- to 64-year-old) men, originally recruited in 1986 (Miller *et al.* 1995). Nine general practices participated in the study, and all patients were free of a history of unstable angina, MI or evidence of silent infarction, coronary surgery, aspirin or anticoagulant therapy, cardiovascular disease, or malignant disease (except skin cancer other than melanoma) at the time of recruitment. Each participant was instructed to avoid a cooked breakfast or a heavy meal before examination. They had refrained from smoking and vigorous exercise from midnight of the previous night. Baseline characteristics and demographic information were ascertained by means of a questionnaire completed at the beginning of the study. Every year patients who were still alive were recalled for interview so that coronary events could be identified, and lipid measurements were repeated. A routine ECG was done at the sixth examination. The endpoints were fatal coronary heart disease events and non-fatal MI, coronary artery surgery and silent MI on follow-up ECG. At present, 227 CHD events have been recorded, with a median length of follow-up of 10.6 years (9 days to 13.7 years). Genotyping was carried out on 2,676 men; the rest were non-Europeans or had incomplete measurements and were therefore ineligible.

3.2.2 WOSCOPS cohort

The West of Scotland Coronary Prevention Study consists of 6,595 moderately hypercholesterolemic men aged 45 to 64 years with a mean baseline total serum/plasma cholesterol of 7.0 mmol/l, HDL-cholesterol of 1.2 mmol/l and triglyceride of 1.8 mmol/l, who have been recruited by population screening (WOSCOPS Group 1997). The principal exclusion criteria were previous MI, angina pectoris requiring hospitalisation within the previous 12 months, and life-threatening non-cardiac illness. Self-reported hypertension, diabetes mellitus, history of smoking, medication, employment, marital status, educational achievement, and family history of premature coronary heart disease mortality were noted. 580 individuals who during the course of the 5 year follow-up had experienced a fatal or non-fatal MI, sudden coronary death or required coronary artery bypass graft or angioplasty were defined as cases, and were matched with 2 controls drawn from the cohort. They were matched on the basis of

age (using 2-year age categories) and smoking status. Participants were randomised to either receive 40 mg Pravastatin each evening or a placebo. They were followed up every 3 months for a mean of 4.9 years. End points were identified from information received at routine trial visits, by analysis of annual electrocardiograms and from death reports. Genotyping was carried out on 1,596 men from the placebo group, from which 348 had reported an event.

3.2.3 HIFMECH cohort

The Hypercoagulability and Impaired Fibrinolytic function MECHanisms predisposing to MI (HIFMECH) study is a large European multicentre case-control study, comprising Caucasian, male survivors of a MI aged under 60 years (excluding patients with familial hypercholesterolaemia and insulin-dependent diabetes mellitus), and age-matched controls from the same regional areas as cases (Juhan-Vague *et al.* 2002). Men were recruited from four centres: Stockholm (Sweden) and London (UK) representing the North and Marseilles (France) and San Giovanni Rotondo (Italy) representing the south of Europe. Patients and controls were investigated in parallel during the early morning following an overnight fast, when blood samples were obtained. Patients were recalled for assessment between 3 and 6 months after the MI. At the time of the present study, 491 patients and 517 controls were available for genotyping.

3.2.4 Army volunteer exercise study

This cohort comprised 250 healthy Caucasian individuals from the Army Training Regiment, Bassingbourn, UK who were initially recruited at the beginning of an 11-week period of basic training, when venous blood samples were drawn from the antecubital vein (Brull *et al.* 2003). The end of the training was marked with an intensive 48-hour final military endurance exercise (FME), which has been shown to induce an acute inflammatory response. Blood samples were then taken on three occasions (2, 48 and 96 hours) following return from the FME. DNA was extracted from all individuals for genotyping and analysis.

3.2.5 Electron Beam Computerised Tomography (EBCT) Study

EBCT was used to compare coronary artery calcification and coronary risk factors in Type 1 diabetic patients and non-diabetic participants A random sample of Type 1 diabetic men (n=104) and women (n=95) aged 30 to 55 years was taken from the diabetes registers of five London hospitals (Colhoun *et al.* 2000). Type 1 diabetes was defined by age of onset ≤25 years and insulin treatment within one year of diagnosis. A random sample of the general population (94 men and 107 women), stratified to have a similar age and gender distribution to the patients with diabetes, was drawn from the lists of two London general practices. Subjects were included regardless of any history of heart disease, but only one subject (a diabetic female) had a prior history of any CHD (angina). Patients attended after an overnight fast and had blood samples

taken. A questionnaire was completed and height, weight, waist:hip ratio and blood pressure were measured using standard methods. An Ultrafast CT scanner (IMATRON C-150XL, Imatron San Francisco, US) was used to measure coronary calcification. Two sets of 20 transverse tomograms of 3-mm thickness were taken from the lower margin of the bifurcation of the right branch of the pulmonary artery to the apex of the heart during two breaths. Participants were followed up one year after the initial testing. For the present study only 74 male controls with CRP levels were included.

3.2.6 LEADER cohort

The Lower Extremity Arterial Disease Event Reduction (LEADER) trial is a double-blind randomised controlled trial, carried out in men with lower extremity arterial disease (LEAD) (Meade *et al.* 2002). 1,568 individuals were recruited from 85 practices throughout the UK and 9 hospital vascular clinics, investigating the effects of Bezafibrate in the prevention of CAD events in men with LEAD. Individuals with a previous MI or stroke were eligible for the trial provided that their general management was stable. Active treatment consisted of Bezafibrate 400 mg/day for men with creatinine levels below 135 μmol/l and was placebo controlled and double-blind, with all tablets identical in appearance. Men with entry creatinine levels between 135 and 149 μmol/l took 400 mg on alternate days. Participants were seen 1 month after treatment started and then at 3 month intervals. Blood was taken for measurements twice at baseline, once at 1 and 3 months and then once at six monthly intervals. Genotyping and analysis was carried out on 647 Caucasian men in active treatment and 419 on placebo.

3.2.7 UDACS cohort

The UCL Diabetes and Cardiovascular disease Study (UDACS) study was a cross sectional case-control study in design aimed to evaluate risk factors for coronary heart disease on 1020 subjects with diabetes mellitus (DM) (Dhamrait *et al.* 2004). Patients were recruited from the diabetes clinic at University College London Hospitals NHS Trust (UCLH) between the years 2001–2. All patients had diabetes according to WHO criteria. No subjects requiring renal dialysis were recruited. Analyses were confined to Caucasian subjects with Type 2 DM. Subjects were categorised by the presence/absence of clinically manifest CVD. CVD was recorded if a patient had one or more of CHD, peripheral vascular disease (PVD) or cerebrovascular disease (CbVD). The presence of CHD was recorded if any patient had positive coronary angiography or angioplasty, coronary artery bypass, a positive cardiac thallium scan or exercise tolerance test, documented evidence of myocardial infarction or symptomatic/treated angina. The presence of PVD was recorded in any patient with absent peripheral pulses and abnormal lower limb doppler pressures or an abnormal lower limb angiogram, previous angioplasty or limb by-pass graft. CbVD was recorded if a patient had been investigated for symptoms or signs consistent with a cerebrovascular accident and had a

brain computed tomography (CT) scan showing any evidence of infarction (diffuse/localised) or haemorrhage. Subjects who were asymptomatic for CHD/CbVD/PVD or had negative investigations were categorised as having no CVD. For the current study 348 Caucasian males without coronary heart disease were included.

3.2.8 Ely Study

The Ely Study is a prospective population-based cohort study of the aetiology and pathogenesis of type 2 diabetes and related metabolic disorders (Wareham *et al.* 1999). The 1122 volunteers were originally recruited in 1990–1992 from a population-based sampling frame consisting of all people in Ely aged between 40 and 65 years in the single GP practice in Ely, Cambridgeshire. At baseline, socio-economic status (SES) was collected by self-report. Participants were restudied in 2001–2003 (follow-up), representing a median ± SD follow-up period of 5.6 ± 0.30 years. Complete data on biochemical and anthropometric variables were available in 839 subjects. The procedure for data collection was the same at baseline and follow-up. Participants attended the laboratory after an overnight fast. Height and weight were measured with the participant wearing light indoor clothing and BMI was calculated. Waist and hip circumference were measured in duplicate using a metal tape. Diastolic and systolic blood pressure was recorded three times using an Accutorr automatic sphygmomanometer (Datascope, Cambridge, UK). Blood samples were taken at fasting and 30 and 120 min after a 75-g oral glucose load. All samples were permanently stored at –70°C within 4 hours. Plasma glucose was measured in the routine UK National Health Service laboratory at Addenbrooke's Hospital by the hexokinase method, and total serum cholesterol, HDL, and triglyceride were measured using the RA 1000 (Bayer Diagnostics, Basingstoke, Hants, UK). Genotyping was conducted in 600 subjects in Cambridge, with the collaboration of Dr Manjinder Sandhu and Professor Nick Wareham.

3.2.9 Periodontal disease cohort

This study was conducted on healthy individuals of mixed ethnicity referred to the Department of Periodontology of the Eastman Dental Hospital, UCL, UK (D'Aiuto *et al.* 2005). Only subjects presenting with severe (probing pocket depths greater than 6 mm and marginal alveolar bone loss greater than 30%), and generalized (at least 50% of teeth affected) periodontitis were invited to participate in the study. Exclusion criteria included known systemic diseases (hypertension, diabetes, dyslipidemia and history of myocardial infarction or stroke), history and/or presence of other acute or chronic infections, systemic antibiotic treatment in the preceding three months, treatment with any medication known to affect the serum level of inflammatory markers (e.g. statins, steroids, hormone replacement therapy) and or pregnant or lactating females. A baseline visit was conducted for clinical staff to collect a complete medical history and standard clinical periodontal parameters (periodontal probing

depths, recession of the gingival margin and clinical attachment levels). Body mass index (BMI) was calculated as the weight in kilograms divided by square of the height in metres. Subjects thereafter received either an intensive session of subgingival mechanical instrumentation under local anaesthesia (within a 6 hour period) or standard periodontal treatment of scraping by clinical staff at the department of Periodontology at the Eastman Dental Hospital. Instrumentation was performed using a piezoelectric instrument (EMS, Nyon, Switzerland) equipped with fine tips for access in the subgingival environment. Extraction of compromised teeth was performed during the same session according to standard clinical practice (D'Aiuto *et al.* 2005). Periodontal outcomes were re-assessed for 1 month following completion of periodontal therapy. During the study period, patients remained stable and there were no changes in lifestyle including exercise, diet, smoking or medications. Serial blood samples were obtained at baseline, one, seven and thirty days after periodontal therapy, from which DNA was extracted for genotyping. An original set of 55 participants was recruited for this study on which intensive periodontal treatment was carried out (data published), following which another 134 patients were additionally recruited for a randomised clinical trial where half the patients received the intensive periodontal treatment and the other half received standard treatment. Only patients receiving the intensive treatment were included in the analysis. In addition, 45 patients were later recruited who also received intensive periodontal treatment and fulfilled the inclusion criteria for the study.

3.2.10 ETNIAS cohort

The ETNIAS study is a population-based cross-sectional survey, using a random sampling scheme composed of a three stage sampling process, where one person from each selected household was randomly chosen. Seven Colombian regions with different ethnic backgrounds were selected (Bucaramanga, Marinilla, Pereira, Pasto, Neiva, Quibdó and Cartagena). A total of 1028 non-related adult subjects of White-Hispanic, Afro-Caribbean, indigenous Amerindian, or mixed ethnic background were selected. The Amerindian reservations (Tama Amerindians and Emberá Amerindians) were located in two rural areas, where inhabitants do not speak Spanish and have continued within their own cultural habits. Inclusion criteria for the study included being born and living all their life in the region they were sampled, and having previous relatives (parents and grandparents) also born in the same region. All included participants answered a survey in regards to their age, gender, socio-economic position, ethnic background, and family history of myocardial infarction, stroke, diabetes and arterial hypertension. Blood samples were also taken at the time of the survey and were transported into a conventional freezer at 4 °C within 6 hours from each region to the central laboratory at Universidad Autonoma de Bucaramanga (UNAB) in Bucaramanga for analysis. DNA was extracted from all individuals in Columbia and transported to UCL for genotyping. The study received ethical approval both in Columbia and London.

Table 3.1. Study demographics and CRP measures.

Study	Study Design & Median follow-up (years)	Study population	CRP assay*	CRP measure	Mean CRP (mg/L) (SD)
NPHSII	Prospective Cohort, (10.6)	Healthy middle-aged Caucasian men.	ELISA	Plasma	2.46 (2.51)
LEADER	Nested case-control study from a clinical trial of fibrate treatment, (4.6)	Caucasian men with lower extremity arterial disease.	ELISA	Plasma	
WOSCOPS	Nested case-control in a clinical trial of pravastatin treatment, (4.9)	Moderately hypercholesterolemic Caucasian men.	ELISA	Plasma	Cases: 3.9 (4.6) Controls: 3.3 (4.45)
HIFMECH	Case-control, (N/A)	Cases: male MI survivors. Controls: men matched by age and regional areas.	ELISA	Plasma	Cases: 2.37 (2.51) Controls: 1.45 (1.41)
Army	Cross-sectional, (N/A)	Healthy Caucasian Army recruits.	Immunonephelometry	Plasma	0.59 (0.04)
UDACS	Case-control, (N/A)	Cases: men and women with DM and CVD. Controls: men and women with DM but without CVD.	ELISA	Plasma	Cases: 1.77 (1.59) Controls: 1.66 (1.42)
EBCT	Case-control (N/A)	Cases: men and women with type-1 DM. Controls: healthy men and women matched by age and gender.	ELISA	Plasma	
Ely	Prospective Cohort, (5.6)	Healthy middle-aged Caucasian men and women.	Immunoturbidometric assay	Plasma	2.32 (6.32)
Periodontal	Prospective trial following treatment (0.08)	Healthy subjects of mixed ethnicity with severe Periodontal disease requiring intensive treatment.	Immunoturbidometric assay	Serum	3.06 (3.65)
ETNIAS	Cross-sectional, (N/A)	Healthy subjects born and living in particular region of particular ethnic background.	Immunoturbidometric assay	Plasma	1.21 (2.57)

*All assays were calibrated against an international reference standard (Pearson *et al.* 2003a).

Table 3.2. Genotyping in studies evaluated.

Study	Total sample size	Number genotyped	Main exclusion criteria
NPHSII	3012	2676	Pre-existing CVD, Coronary surgery, Aspirin or anticoagulant therapy, Malignant disease, No CRP measures.
LEADER	1568	1066	Unstable angina, Total cholesterol <3.5 or >8.0 mmol/l., Significant renal or hepatic disease or malignant disease, No CRP measures.
WOSCOPS	6595	1451	MI, or angina pectoris requiring hospitalisation, Life-threatening non-cardiac illness, No CRP measures.
HIFMECH	Cases: 533 Controls: 575	Cases: 491 Controls: 517	FH, Insulin-dependent DM, No CRP measures.
Army	250	219	No CRP measures.
UDACS	Controls: 449	348	No CRP measures.
EBCT	Controls: 94	72	No CRP measures.
Ely	1122	598	Malignant disease, DM, No CRP measures.
Periodontal	231	164	Hypertension, DM, MI or stroke, antibiotic treatment, statin or HRT therapy, No CRP measures.
ETNIAS	1028	1028	More than one person from a selected household.

3.3 DNA extraction from blood samples

DNA extraction from whole blood samples was either performed via a salting-out method or using a QIAamp DNA blood minikit (QIAGEN).

3.3.1 DNA extraction using salting-out method

10ml of a solution of 109.54g sucrose dissolved in 10ml of Tris pH7.5 (1M), 5ml of $MgCl_2$ (1M) and 10ml of Triton-X-100 made up to 1L with deionised water (reagent A) was used to lyse the cells by mixing and centrifugation. The supernatant was discarded and the pellet was resuspended in 10ml of reagent A. This step was repeated to increase the purity of the pellet. 2ml of a solution of 2.34g NaCl in 1ml Tris-HCl (1M), 0.4ml Na-EDTA (0.5M), 90ml deionised water and 10ml of 10% SDS (reagent B) was added to the pellet to complete the nuclear lysis step, followed by addition of 1ml of sodium perchlorate (5M) for deproteinisation. 2ml of chloroform was added and thoroughly mixed, followed by centrifugation to allow separation of protein. The upper aqueous phase containing the DNA was transferred to a sterile propylene tube and the lower phase discarded. 100ml of ice-cooled ethanol was added to allow the DNA to precipitate. The DNA was removed from the tube by inserting and swirling a sterile Pasteur pipette into the solution, upon which it adheres, and washing the pipette in 70% ethanol to remove salt and organic molecules, followed by resuspension in TE buffer (Tris-HCl and EDTA).

3.3.2 DNA extraction using QIAamp DNA blood minikit (QIAGEN©)

DNA extraction was performed using the QIAamp DNA blood minikit protocol. Cell lysis was achieved by adding proteinase K to 200µl of whole blood and incubating at 56° for 10 minutes. The lysate was then pipetted into a 1.5ml spin column containing a silica-gel membrane filter. Buffers were added to the lysate containing guanidinium chloride, which disrupts the bonds between water molecules and DNA and makes the DNA adsorb to the silica membrane filter. The membrane was then washed with a series of high-salt alcohol solutions and centrifuged briefly to allow protein and other contaminants to pass through the filter and be removed, leaving purified DNA bound to the membrane. Purified genomic DNA was eluted from the membrane using an elution buffer containing 10mM Tris-Cl and 0.5mM EDTA (pH 9.0). An average of 6-8µg of DNA is extracted from 200µl of blood using this method.

3.4 Genotyping polymorphisms in the CRP gene

Polymorphisms in the CRP gene were genotyped by RFLP analysis, TaqMan assays or pyrosequencing. Genotyping was conducted in 96 well or 384 well plates. Each plate contained at least two negative controls (no DNA), and repeat samples were present on each plate. In addition, where polymorphisms had low minor allele frequencies, a minimum of three positive controls for the rare homozygous alleles were also present on each plate.

3.4.1 Genotyping by RFLP analysis

Genotyping of the +1444C/T polymorphism (rs1130864) was carried out using the restriction endonuclease *SduI* that recognises the common allele, but not the rare allele. The forward primer sequence was mismatched to force an allele-specific restriction enzyme site into the PCR product. The region containing the relevant polymorphism was amplified by PCR as shown in Figure 3.1.

Figure 3.1. Region of CRP gene (181bp) being amplified to genotype the +1444C/T polymorphism (highlighted in green). The arrows show the start of the primers (underlined). The +1444C/T polymorphism is highlighted in yellow.

Forward primer: AGC TCG TTA ACT ATG CTG GGG CA

Reverse primer: CTT CTC AGC TCT TGC CTT ATG AGT

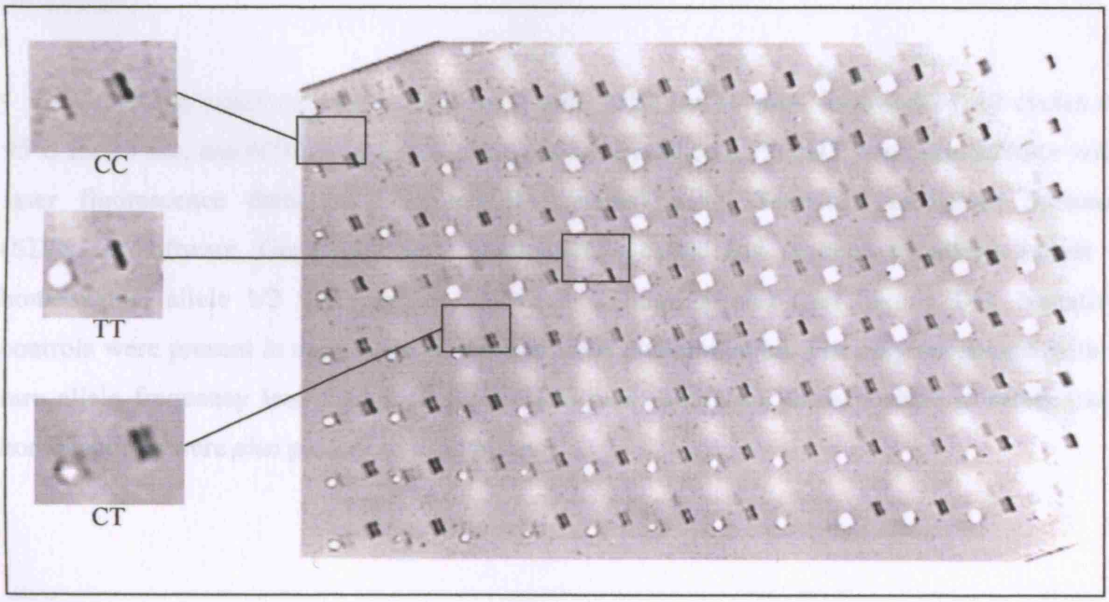
```
1380  CCTCAGCGCC TGAGAATGGA GGTAAAGTGT CTGGTCTGGG AGCTCGTTAA
1430  CTATGCTGGG AAACGGTCCA AAAGAATCAG AATTTGAGGT GTTTTGTTTT
1480  CATTTTTATT TCAAGTTGGA CAGATCTTGG AGATAATTTC TTACCTCACA
1530  TAGATGAGAA AACTAACACC CAGAAAGGAG AAATGATGTT ATAAAAAACT
1580  CATAAGGCAA GAGCTGAGAA GGAAGCGCTG ATCTTCTATT TAATTCCCCA
```

DNA samples were heated at 80°C for 10 min to dry them, before the PCR reaction mix was added. PCR was carried out in a 20µl reaction volume containing 2.5µl of 5ng DNA, 10x Polmix buffer containing 500mM KCl, 100mM Tris (pH 8.3), 0.01% Gelatin, and 2mM of each dNTP, 3.5mM $MgCl_2$, 6pmol/µl of each oligonucleotide primer and 0.4U *Taq* polymerase. Reactions were carried out in 96-well microtitre plates (ABGene). Cycling conditions were

95°C for 4 min, followed by 35 cycles of 95°C for 30 sec, 55°C for 30 sec and 72°C for 30 sec, and a final extension at 72°C for 3 min, on a Techne-Genius 96-well PCR machine. A negative control was present in each plate to test for DNA contamination.

Restriction enzyme digestion was carried out in a 10µl reaction volume containing 5µl PCR product, 10 units/µl *SduI* restriction enzyme and its buffer (10x) containing 10mM Tris-HCl, 3mM $MgCl_2$, 150mM NaCl, and 0.1mg/ml BSA, and left overnight at 37°C. The *SduI* restriction enzyme cleaves the 181bp PCR product into 23bp and 158bp fragments only in the presence of the common C allele. Digestion products were electrophoresed at 100V for 1 hour on 7.5% MADGE (Microtitre Array Diagonal Gel Electrophoresis) gels containing 10x TBE, 30% polyacrylamide (acrylamide/bisacrylamide, 19:1) and TEMED (N,N,N'N'-Tetramethyl ethylene- diamine), and visualised by ethidium bromide fluorescence under U.V. light.

Figure 3.2. MADGE (Microtitre Array Diagonal Gel Electrophoresis) gel showing +1444 CC and TT homozygous and CT heterozygous subjects. The *Sdu*I restriction enzyme has cleaved the 181bp PCR product into 23bp and 158bp fragments in the presence of the C-allele.



3.4.2 Genotyping using TaqMan assays

Genotyping for the CRP –717A/G, -305G/A, +1444C/T, +2302G/A and +4899T/G polymorphisms was carried out using primers and probes (see Table 3.3) designed by Applied Biosystems (ABI TaqMan assays by design) in 384-well Thermo-fast PCR plates (ABGene). DNA samples were left to dry for a minimum of a day before the PCR reaction mix was added. PCR was carried out in a 9µl reaction volume containing 4µl of 1.25ng DNA, 2x TaqMan

Universal PCR master mix containing AmpliTaq Gold DNA Polymerase, AmpErase® uracil-N-glycosylase (UNG), dNTPs with dUTP, and a dichlororhodamine acceptor dye (ROX) passive reference to which samples are normalized (Applied Biosystems), 80x Assay mix containing the forward and reverse primers and VIC and FAM labelled probes, and Sigma water to dilute the genomic DNA.

The TaqMan probe contained a reporter dye at the 5' end and a quencher dye at the 3' end (see Figure 3.3). During PCR, if the target allele of interest was present, the probe specifically annealed between the forward and reverse primer sites. The 5'-3' nucleolytic activity of the DNA Polymerase cleaved the probe between the reporter and the quencher only if the probe hybridised to the target. Cleavage of the probe separated the reporter dye and the quencher dye, resulting in increased fluorescence of the reporter. The probe fragments were then displaced from the target, and polymerisation of the strand continued. Accumulation of PCR products was detected directly by monitoring the increase in fluorescence of the reporter dye. When the probe was intact (i.e. when the correct variant for hybridisation was not present), the reporter dye remained near the quencher dye and resulted in suppression of the reporter fluorescence.

Cycling conditions were 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec, and 60°C for 1 min. Genotypes were read on a 7900HT sequence detector with laser fluorescence detection (Applied Biosystems) using Sequence Detection Systems (SDSv2.1) software. Genotypes were automatically called and grouped by alleles (allele 1 homozygote, allele 1/2 heterozygote, or allele 2 homozygote) (see Figure 3.4). Negative controls were present in each plate to test for DNA contamination. For polymorphisms with a rare allele frequency less than 0.10 (+4899T/G and –305G/A), three positive controls (rare homozygotes) were also present in each plate.

Figure 3.3. Principles of TaqMan genotyping. Probes are labelled with Vic and Fam fluorescence (ABI handbook, http://www.appliedbiosystems.com).

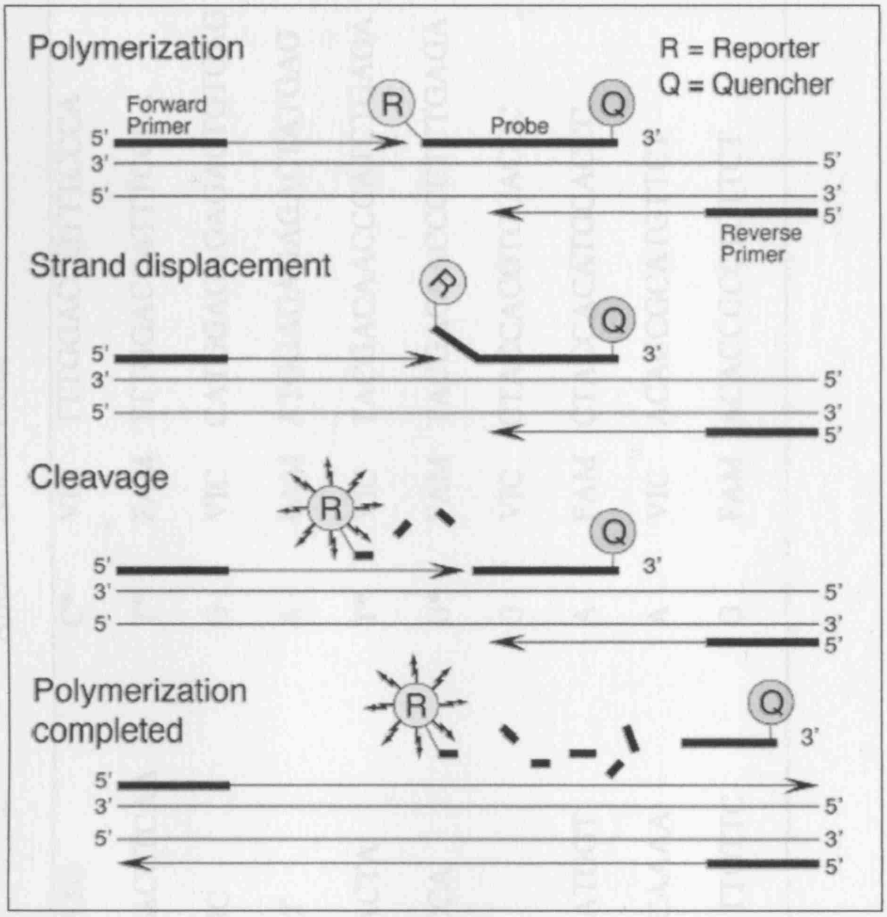
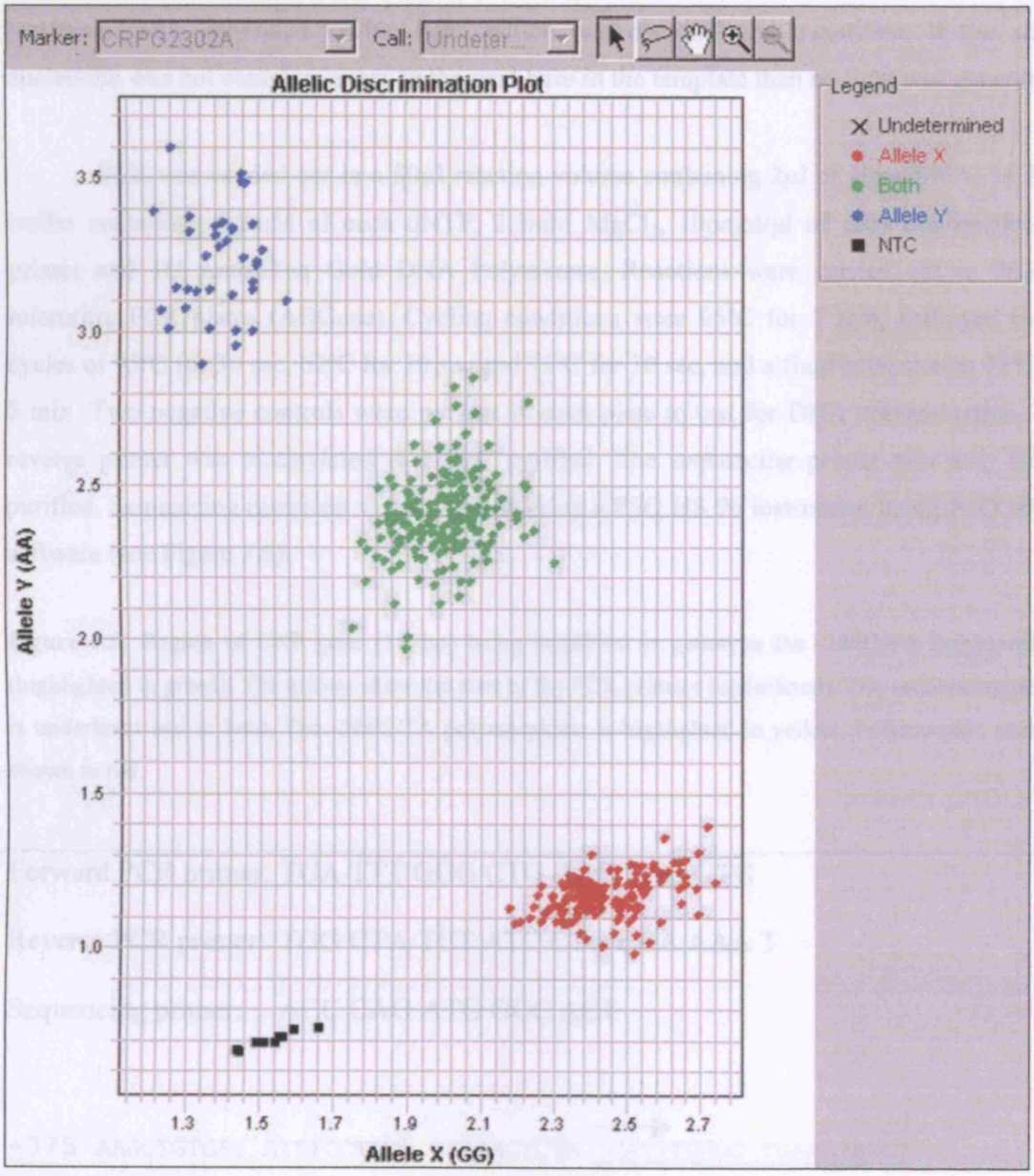


Table 3.3. Primers and probes used for genotyping the CRP polymorphisms. The polymorphic sites are highlighted in yellow in the probe sequences.

SNP	Primer (5'-3')		Allele	Allelic Probe	
+1444C/T	Forward	GGTCTGGGAGCTCGTTAACTATG	C*	VIC	TTTGGACCGTTTCCCA
	Reverse	TCCAACTTGAAAAACAAAACACCTCAA	T*	FAM	TTTGGACCATTTCCCA
+2302G/A	Forward	CACCAGTAGCCATCTTGTTTGC	G	VIC	CATGGAGAGAGACTGTGAG
	Reverse	CCACTTCCAGTTTGGCTTCTGT	A	FAM	ATGGAGAGAGACTATGAG
+4899T/G	Forward	TTATCCTAGGACAACTGCCCACTA	T*	VIC	TACGACAACCCATCTGAGA
	Reverse	GGAGCTGAAGAGAAGGAATCCA	G*	FAM	TACGACAACCCCTCTGAGA
-305G/A	Forward	GGGCTGAAGTAGGTGTTGGA	G	VIC	CTACCACGTGCACCC
	Reverse	TCCTGCGAAAATAATGGGAAATGGT	A	FAM	CTACCACATGCACCC
-717A/G	Forward	GCTGAGAAAATGTGTCCATGCAAAA	A	VIC	ACACCGCATGTTCT
	Reverse	TCCTGTGTCCAAGTATTCTCATTGTTC	G	FAM	ACACCGCGTGTTCT

* probes are designed to the reverse strand.

Figure 3.4. Plot of +2302G/A genotypes generated from TaqMan assay.



+2302GG genotypes are shown in red, +2302GA genotypes are shown in green and +2302AA genotypes are shown in blue.

3.4.3 Genotyping by pyrosequencing

Genotyping for the CRP –286C/T/A polymorphism (rs3091244) was by pyrosequencing for real time sequencing in the NPHSII, Periodontal disease and ETNIAS cohorts following a standard Biotage AB protocol (www.biotage.com). The region containing the –286C/T/A polymorphism was amplified by PCR as shown in Figure 3.5. Biotinylated PCR products were converted to single stranded templates onto which a sequencing primer was

annealed. Addition of dNTPs to the template released pyrophosphate that is used in a reaction to produce ATP, which subsequently drives the conversion of luciferin to oxyluciferin, by luciferase. This generated visible light, which was detected and quantified. If the added nucleotide was not complementary to the next base in the template then no light was generated.

PCR was carried out in a 50µl reaction volume containing 2µl of 10ng DNA, 1x PCR buffer containing 0.2mM of each dNTP, 2.5mM $MgCl_2$, 10pmol/µl of each oligonucleotide primer and 1U AmpliTaq Gold DNA Polymerase. Reactions were carried out in 96-well microtitre PCR plates (ABGene). Cycling conditions were 95°C for 7 min, followed by 50 cycles of 95°C for 30 sec, 62°C for 30 sec and 72°C for 30 sec, and a final extension at 72°C for 5 min. Two negative controls were present in each plate to test for DNA contamination. The reverse primer was biotinylated & HPLC purified. The sequencing primer was also HPLC purified. Sequencing reactions were analysed using a PSQ HS 96 instrument using PSQ HS 96 software (see Figure 3.6).

Figure 3.5. Region of CRP gene (107bp) being amplified to genotype the -286C/TA polymorphism (highlighted in green). The arrows show the start of the PCR primers (underlined). The sequencing primer is underlined and in bold. The -286C/TA polymorphism is highlighted in yellow. Polymorphic sites are shown in red.

Forward PCR primer: TGA TTT GGG CTG AAG TAG GTG

Reverse PCR primer: TGG CTA TCT ATC CTG CGA AAA T

Sequencing primer: ACC CAG ATG GCC ACT

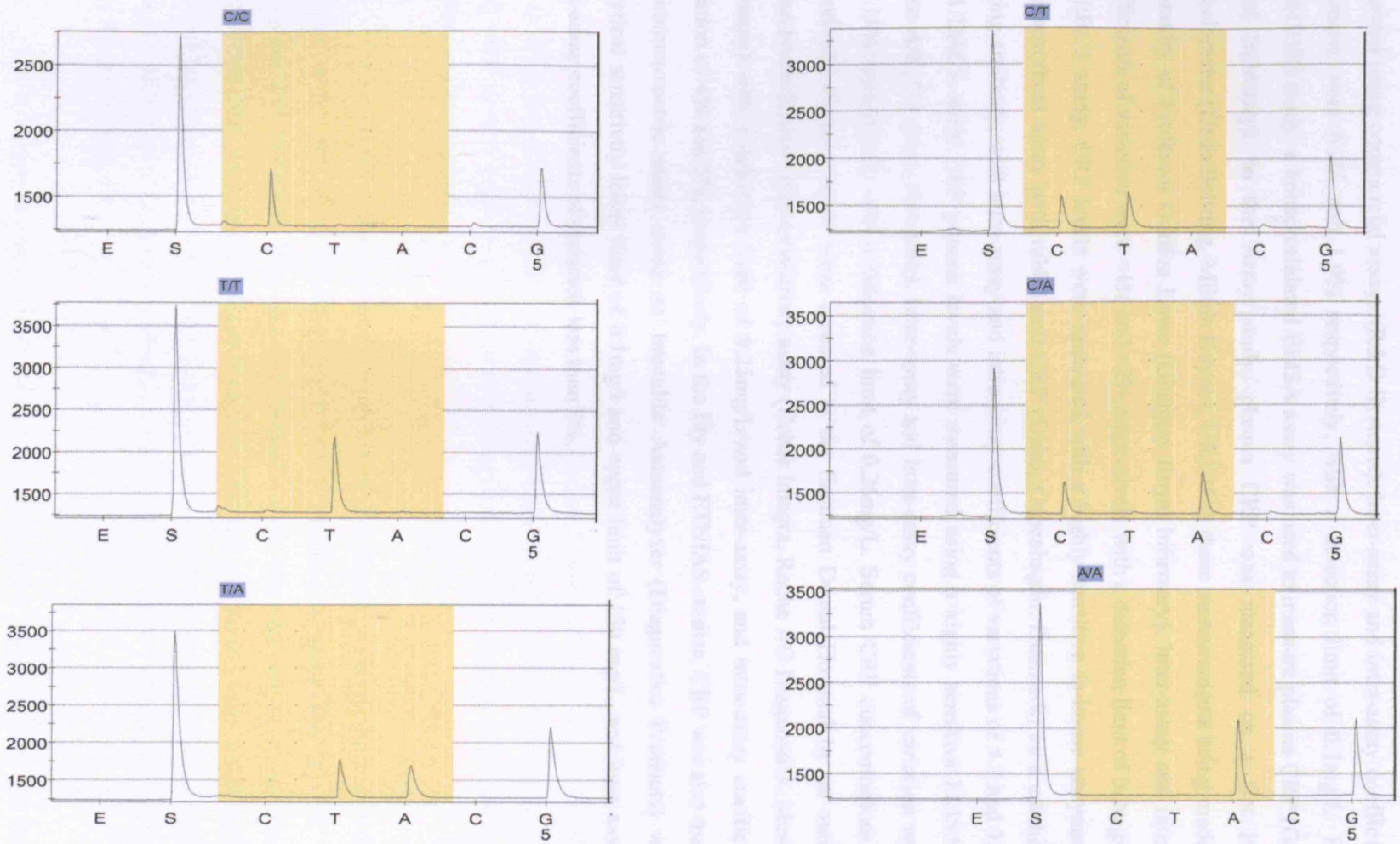
-375 AAGCTGTCAG ATTTCCTTTG TCAAACTCTA TGATTTGGGC TGAAGTAGGT

-325 GTTGGAGAGG CAGCTACCAC GTGC**ACCCAG ATGGCCACTC** GTTTAATATG

-275 TTACCATTTC CCATTATTTT CGCAGGATAG ATAGCCAAAG TGGAGCCCTG

-225 AGAGATTTCT TCATTTTTCC TGTCATAAAG AATTGGTAAT TCAGTAGTCA

Figure 3.6. Visual plots for each genotype generated for the –286C/T/A polymorphism by pyrosequencing.



Peaks within the yellow box indicate the polymorphic site where one peak indicates a homozygous individual and two peaks indicates a heterozygous.

3.5 Measurement of CRP

Plasma CRP concentrations for the NPHSII, LEADER and HIFMECH studies were measured using commercial assays (R&D Systems). Inter-assay and intra-assay coefficients of variations were 6.2% and 1.9% respectively, with a detection limit of 0.1mg/L. For the WOSCOPS study, a house-validated ELISA assay was used to measure plasma CRP (Glasgow Royal Infirmary). In the Army study, plasma CRP was measured on a BN ProSpec nephelometer (Dade Behring, Milton Keynes, UK), with these measurements being made in the laboratory of Professor Gordon Lowe (Glasgow Royal Infirmary). Inter-assay and intra-assay coefficients of variation were <4% and <2% respectively with a detection limit of 0.2mg/L. For the EBCT study, CRP levels were measured with a highly sensitive in-house enzyme-linked immunosorbent assay with rabbit anti-CRP (Dako, Copenhagen, Denmark) as a catching and tagging antibody, with inter-assay and intra-assay coefficients of variations of 4.7 and 3.8%. In the UDACS study CRP plasma levels were measured using a highly sensitive ELISA assay (Dako A/S, Glostrup, Denmark). Inter-assay and intra-assay coefficients of variation were 8% and 10% respectively with a detection limit of 0.26mg/L. Serum CRP concentrations in the Periodontal disease cohort were assessed at the Eastman Dental Hospital by an automated immunoturbidimetric high-sensitivity assay (Cobas Integra, Roche AG Diagnostics, Mannheim, Germany) with a detection limit of 0.25mg/L and inter-assay, and intra-assay coefficient of variation of 4% and 5%, respectively. In the Ely and ETNIAS studies, CRP was also measured by immunometric assay using an Immulite Autoanalyzer (Diagnostics Products) with an analytical sensitivity lower limit of 0.1mg/l and upper limit of 150 mg/L, and inter-assay and intra-assay coefficients of variation less than 8%.

3.6 Statistical analyses

For general analyses, log-transformations were conducted for data not normally distributed to generate geometric means and approximate standard deviations. The association of CRP with other risk factors was assessed by Pearson correlation using measurements made in the same year as the CRP measure where possible. Survival analysis was carried out using Cox proportional hazards models, with events occurring before the CRP measurement being excluded from the analysis, and the time of the CRP measurement being taken as the start of follow-up. To allow for differences according to age and practice, age was included as a covariate in the model and the data stratified by practice (using the strata option in STATA). The performance of CRP as a screening test for CHD was assessed by means of disease detection and false positive rates, and by using the area under the receiver-operating characteristic (ROC) curve as a discriminatory test. Means and standard deviations of log-transformed CRP, as well as certain other risk factors and biomarkers in those with and without CHD were used to construct normal curves to determine the overlap between the distributions in the two groups.

For meta-analyses, a random effects model was used, in order to allow for any heterogeneity across studies. The geometric mean was used wherever possible, or calculated from the arithmetic mean. However, some studies only used median CRP values and it was not possible to obtain geometric mean values. The majority of studies reported standard deviation (SD) values, although where these values were not available, the SD from the largest study was used. In cases where the SD values were implausible due to small studies sizes, the SD from the largest study was also used. The DerSimonian and Laird Q test, and the I^2 test (Higgins *et al.* 2003) were used to evaluate the degree of heterogeneity between studies, and funnel plots to evaluate small-study bias.

For genotypic analysis, observed numbers of each genotype were compared with those expected under the Hardy-Weinberg equilibrium assumption using the χ^2 test. Allele frequencies were calculated from genotype frequencies for own data and for the meta-analyses of published data. Associations of genotype with CRP and risk factors were tested using one-way analysis of variance (ANOVA) or Kruskal-Wallis tests for continuous variables and χ^2 or Fisher's exact test for categorical variables. In addition, "per allele" effects were determined using regression analysis.

Data were analysed using the Review Manager version 4.2 software from the Cochrane Collaboration 2003, Stata 8.0 (Stata Corporation, College Station, Texas, 2003), SPSS version 11 (Chicago, IL), and Comprehensive Meta-analysis (CMA) version 2.0 software (Biostat).

Haplotype analysis was performed using a maximum likelihood model based on the stochastic-EM algorithm implemented in the THESIAS program (http://www.genecanvass.org). THESIAS allows the simultaneous estimation of haplotype frequencies and of their associated effects on the phenotype of interest. The haplotype combining the most frequent alleles at each locus was used as the reference. A global p value was calculated using differences in log-likelihood assuming an additive model of haplotype effects.

The linkage disequilibrium was measured between polymorphisms in the CRP gene using the software packages TagIT v.3 (http://www.genome.duke.edu/resources/computation/software), PHASE (v 2.0) (http://pga.gs.washington.edu/) and Tagger (within the Haploview software) (http://www.hapmap.org/). These programs employ a linkage disequilibrium statistic known as the haplotype r^2 statistic that measures the correlation of alleles at two sites taking into account differences in allele frequencies between two loci and is more stringent than the D' measures of LD. D' is obtained from D, the basic pairwise disequilibrium coefficient that corresponds to the difference between the probabilities of observing the alleles independently in the population.

Results

4. Utility of C-reactive protein in coronary risk prediction

4.1 Aim

To evaluate the performance of CRP in the detection of later coronary heart disease events in initially healthy middle-aged men. To compare its performance with other risk factors and biomarkers, and evaluate whether CRP adds useful predictive information over and above traditional risk factors and risk scores.

4.2 Background

Over the past decade much laboratory evidence has emerged that suggests that inflammatory processes, even those remote from the atherosclerotic lesion could mediate the initiation and evolution of the atherosclerotic lesion, as well as contributing to the ultimate development of acute ischaemic syndromes (Ross 1999). C-reactive protein (CRP) is a major acute phase plasma protein synthesised primarily by the liver that displays a marked rise of its serum concentration in response to infection or inflammation. This has led to the established use of CRP in clinical practice as a marker of progress and response to treatment in infective or inflammatory processes such as pneumonia, meningitis and some autoimmune diseases (Povoa *et al.* 2005; Almirall *et al.* 2004; Paradowski *et al.* 1995; Yildirim *et al.* 2004). However, more recently, in keeping with the interest in a potential link between inflammation and atherosclerosis, the relationship between CRP concentration (as a non-specific marker of inflammation) and CHD has been explored.

4.2.1 CRP and acute cardiovascular events

An early series of studies reported associations between CRP and recurrent cardiovascular events in patients presenting with acute coronary syndrome (ACS), in whom CRP concentrations rise to between 10-100 times their baseline values (Berk *et al.* 1990; Haverkate *et al.* 1997; Morrow *et al.* 1998; Biasucci *et al.* 1999b; Biasucci *et al.* 1999a; Biasucci *et al.* 1999d). A meta-analysis of patients with unstable angina found an increased risk of short-term death or recurrent non-fatal MI in subjects with higher CRP (odds ratio 3.44 for top vs. bottom tertile, quartile or quintile of CRP, 95%CI 2.94-4.03) (Galvani *et al.* 2001). A more recent study, with a larger sample size of over 7000 patients with ACS also found higher concentrations of CRP at presentation were associated with increasing mortality rate at 30 days after hospital admission (odds ratio 1.19 per quartile of CRP, p=0.006) (James *et al.* 2003). The prognostic utility of CRP measurement in addition to the Thrombolysis In Myocardial Infarction (TIMI) risk score for ACS and acute MI (a score that incorporates clinical and demographic information) has also been examined using sensitivity and specificity analyses. An

elevated plasma CRP level was found to provide additional prognostic information to the validated TIMI risk score for an endpoint of 30-day mortality with an area under the receiver operating characteristic curve (AUC) value of 0.81 (95% CI 0.79 to 0.85; $p<0.001$) with a cut-off point of 3mg/L (CRP alone had an AUC value of 0.74; $p<0.001$) (Foussas *et al.* 2005) (see section 4.2.5 for an explanation of ROC analysis). The results from these studies suggest that CRP may be useful as a predictive tool and provide additional information in risk stratification in the setting of ACS.

4.2.2 CRP and the longer-term risk of cardiovascular disease

A series of prospective observational studies have shown that higher than average concentrations of CRP within the range normally encountered in health are associated with increased long-term risk of future cardiovascular events 5-10 years later (Ridker & Haughie 1998; Danesh *et al.* 1998; Danesh *et al.* 2000a; Ridker *et al.* 2002; Blake *et al.* 2003). CRP can be measured using widely available commercial and automated assays that detect CRP in the "healthy" range (Pepys & Hirschfield 2003). These assays are precise and reproducible, with the coefficient of variation usually being less than 10% (Pearson *et al.* 2003a). Although there is day-to-day biological variation in CRP values within the same individuals, even in the absence of infection or inflammation, the self-correlation of CRP values is similar to that for cholesterol (Pearson *et al.* 2003b). There also appears to be little evidence for diurnal or seasonal variation (Meier-Ewert *et al.* 2001; Frohlich *et al.* 2002; Pepys & Hirschfield 2003). Coupled with its constant half-life of 19 hours, this means that CRP is a relatively stable analyte in the absence of infection or inflammation, thus fulfilling some of the properties required for an ideal biomarker.

Associations have also been found between CRP concentration and surrogate outcome measures of atherosclerosis such as carotid IMT in cross-sectional studies (Winbeck *et al.* 2002; Sitzer *et al.* 2002; Magyar *et al.* 2003; Cao *et al.* 2003). A meta-analysis carried out by Danesh et al. involving 11 studies that was recently updated to include 22 studies with 2459 cases of CHD and 3969 controls found an odds ratio of 1.45 for coronary disease for individuals in the top tertile of CRP concentration compared to individuals in the bottom tertile (Danesh *et al.* 2004). Associations from individual studies have been more extreme and have led to the proposal that CRP may also be used to help in the prediction of coronary events. Some studies have even suggested that CRP is a stronger predictor of coronary events than LDL cholesterol and may be additive to the Framingham risk score (Ridker 2001a).

4.2.3 Potential clinical application of the CRP-CHD association

While associations between CRP and cardiovascular disease may indicate a causal role for CRP in atherosclerosis (which will be discussed in more detail in Chapter 10), the proposal that measurement of CRP could have utility in cardiovascular disease prediction beyond that

provided by traditional risk factors or scores has generated intense interest. Indeed, CRP need not be causally involved in CHD for it still to serve a useful role in risk prediction. An American Heart Association/Centers for Disease Control scientific statement issued in 2003 concluded that CRP measurement might be used "*at the discretion of the physician, to detect enhanced absolute risk in individuals thought to be at intermediate risk by multiple risk factor scoring (10-year CHD risk in the range of 10% to 20%). However, the benefits of this strategy or any treatment based on this strategy remain uncertain*" (Pearson *et al.* 2003a). The European Society of Cardiology also published their own guidance on the potential role of CRP in CHD prediction in the same year. Their view was more conservative and they stated, "*it is not yet clear how clinicians should use screening results of CRP, and therefore widespread screening is not recommended*" (De Backer *et al.* 2003).

4.2.4 Limitations of CRP measurement in coronary risk stratification

Although numerous studies published in the last few years on CRP and cardiovascular disease risk have shown consistent associations, current opinion on the use of CRP in risk prediction is still divided. Studies by Ridker et al., Koenig et al. and several others arrive at the conclusion that CRP adds prognostic information to the Framingham risk score and that it is a stronger predictor of future cardiovascular events than LDL cholesterol (Ridker *et al.* 2002; Ridker 2003b; Koenig *et al.* 1999; Koenig 2005). Other studies, such as those by Danesh et al. and Wilson et al. suggest that once established risk factors are taken into account, the predictive utility of CRP is limited and that it may not add information to the Framingham risk score (Danesh *et al.* 2004; Wilson *et al.* 2005). This may be because CRP concentration is also associated with many of the established risk factors of CHD including blood pressure, smoking and diabetes that contribute to the Framingham risk score (Folsom *et al.* 2001; Ford *et al.* 2004; Davey Smith *et al.* 2005b).

Although low-grade inflammation likely contributes to some of the observed variation in CRP concentration among healthy subjects, which might account for its association with CHD risk, the precise origin of this inflammation is unclear. The association of CRP with measures of adiposity, and the recognition of adipose as an important source of IL-6 (one of the main stimuli for CRP synthesis), could provide one explanation (Mohamed-Ali *et al.* 1997). However, non-inflammatory stimuli including hormonal, metabolic and other genetic factors, as well as certain orthodox risk factors such as BP and metabolic syndrome may also make an important contribution to the variation in CRP concentration among healthy individuals (Cushman *et al.* 1999b; Barinas-Mitchell *et al.* 2001; Sesso *et al.* 2003; Lakoski *et al.* 2005; Folsom *et al.* 2001; Festa *et al.* 2001). These influences on CRP may work through or independently of inflammatory signals. Because clearance rates of CRP appear rather invariant, observed differences in circulating concentration are likely to reflect alterations in its synthesis

and/or release. Recent molecular studies indicate that CRP transcription is influenced not only by inflammatory cytokines such as IL-6 and IL-1, but also by transcriptional regulators involved in the regulation of lipid and metabolic genes such as USF1 (Corre & Galibert 2005), as well as by endoplasmic reticulum stress pathways that include transcription factors such as CREBH (Zhang *et al.* 2006a). Therefore, there may be reasons for differences in CRP among individuals other than inflammation, which could compromise its predictive utility.

Although associations of CRP concentration with later risk of cardiovascular events are highly reproducible, the size of the associations observed, as summarised in a recent updated meta-analysis of published studies, is much smaller than that previously thought (Danesh *et al.* 2004). The strength of the CRP-coronary event association from this meta-analysis has been shown to be smaller than that observed for risk factors such as blood pressure and cholesterol, which have already been shown to have poor predictive utility for coronary disease when measured in isolation (Wald *et al.* 1999; Jackson *et al.* 2005).

Indeed, much can be learnt about the difficulties of CHD risk prediction by drawing parallels with information on the performance of established risk factors in coronary prediction. For example, for blood pressure and cholesterol, there is a continuous graded relationship with disease risk over the whole range of normal values of these measures, with no threshold below which risk is minimal (MacMahon 2000; Jackson *et al.* 2005). Both of these risk factors are normally distributed within populations with most individuals clustering in the centre of the distribution and fewer individuals at the upper and lower extremes. Therefore, a substantial proportion of coronary events occur among individuals with intermediate levels of blood pressure and cholesterol; almost as many as among those at high risk (Rose 1985; Wald *et al.* 1999). The consequence is that the distribution of risk factors among those who later develop disease and those that remain disease free exhibit substantial overlap, making it difficult to set a threshold level for the risk factor that identifies a large number of disease cases while keeping the misclassification rates low (Rose 1985; Wald *et al.* 1999; Law *et al.* 2004). If the same is true for CRP then it would face similar challenges for disease prediction.

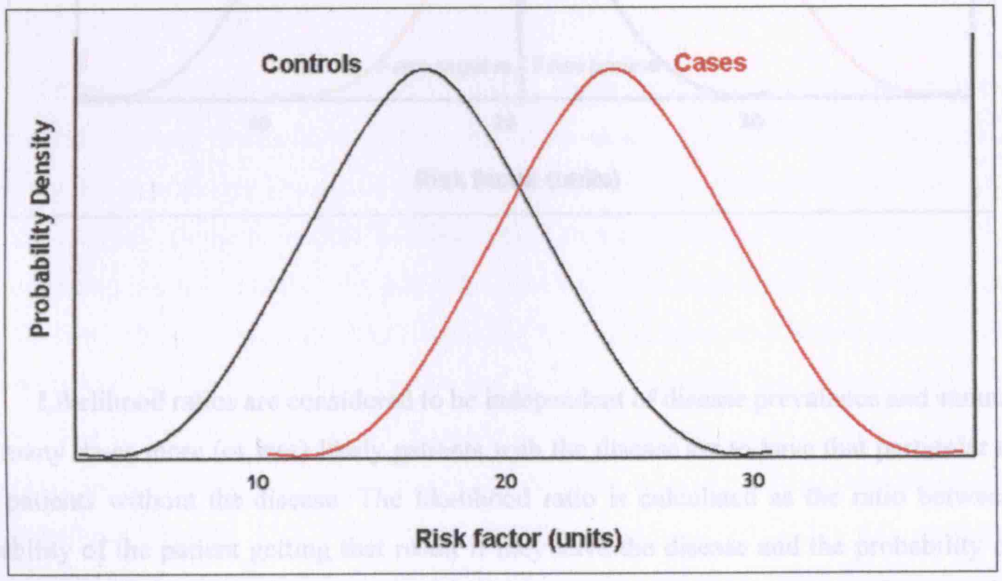
Prior studies on the predictive role of CRP have mostly suggested that it could provide additional information on CHD risk but most of these studies have based this interpretation on the basis of an "independent" association of CRP with disease risk after adjustment for other risk factors with the use of odds ratios (OR) and relative risks (RR), to describe the strength of the association, and on which to base assessment of the performance of CRP as a screening test (Kuller *et al.* 1996; Ridker *et al.* 1997; Koenig *et al.* 1999; Ridker *et al.* 2000a; Ridker & Haughie 1998; Ridker *et al.* 2002; Roivainen *et al.* 2000; Mendall *et al.* 2000; Ridker 2003a; Lowe *et al.* 2004; Koenig *et al.* 2004; Boekholdt *et al.* 2005). However, relative measures of

risk may not be the most appropriate way for evaluating the performance of a biomarker for disease screening. When assessing the utility of the biomarker as a screening test, the critical question is: given a particular value of the biomarker, what is the probability of a future cardiovascular event? (Wald *et al.* 1999; Pepe *et al.* 2004). Sensitivity, specificity, positive predictive values, negative predictive values and likelihood ratios have all been used to assess screening test performance, as they encompass some information not only on the relationship between biomarker and risk, but also on the distribution of the biomarker among later cases and controls (Altman & Bland 1994b; Deeks & Altman 2004). These issues are now elaborated in more detail.

4.2.5 Statistical measures to determine the predictive utility of a marker

Diagnostic or screening tests are used to identify individuals who have, or predict individuals who will develop a disease outcome of interest. For many screening tests, a range of values on a continuous scale is possible, and the performance of a particular threshold value of the test can be assessed using a variety of appropriate statistical measures (Figure 4.1).

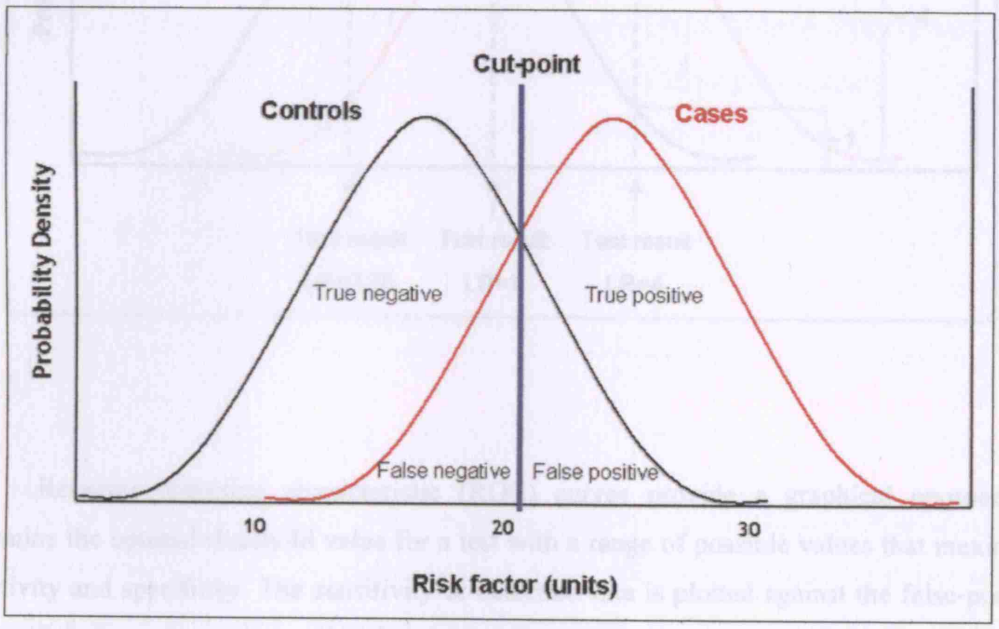
Figure 4.1. Distribution of a continuous risk factor among cases and controls for disease.



These appropriate measures include sensitivity (or detection rate), specificity, false positive rate (1-specificity), positive and negative predictive values and positive and negative likelihood ratios. The sensitivity is the proportion of cases that are correctly identified by the test (true positive divided by the true positive plus false negative) and the specificity is the proportion of controls that are correctly identified by the test (true negative divided by the true negative plus false positive) (see Figure 4.2) (Koepsell & Connell 1985; Altman & Bland

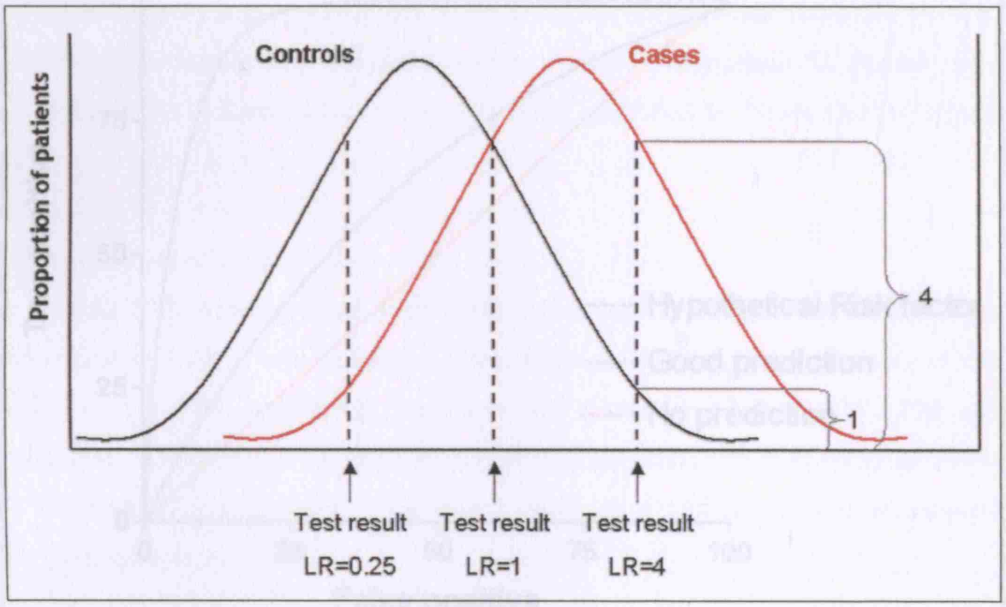
1994a; Schulzer 1994). The positive predictive value is the proportion of individuals with positive test results who are correctly diagnosed (true positive divided by the true positive plus false positive), and the negative predictive value is the proportion of individuals with negative test results who are correctly diagnosed (true negative divided by the true negative plus false negative) (Altman & Bland 1994b).

Figure 4.2. Distribution of a risk factor showing true positive, true negative, false positive and false negative cases and controls that are used to determine sensitivity, specificity, and positive and negative predictive values.



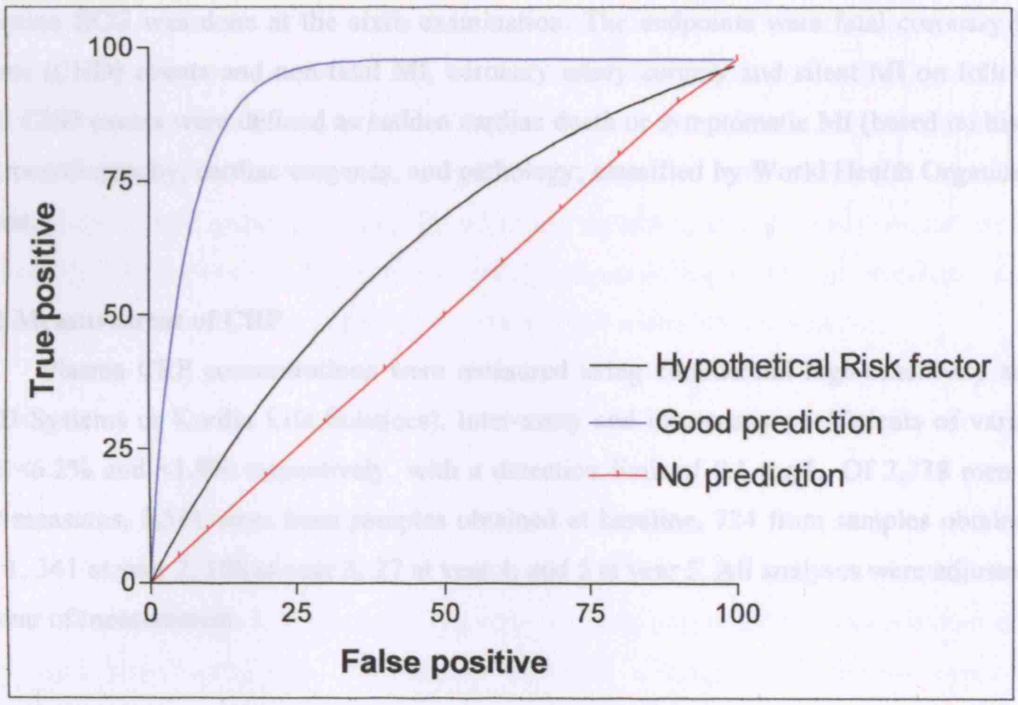
Likelihood ratios are considered to be independent of disease prevalence and summarise how many times more (or less) likely patients with the disease are to have that particular result than patients without the disease. The likelihood ratio is calculated as the ratio between the probability of the patient getting that result if they have the disease and the probability of the patient getting that result if they do not have the disease (see Figure 4.3). A likelihood ratio greater than 1 indicates that the test result is associated with the presence of the disease, whereas a likelihood ratio less than 1 indicates that the test result is associated with the absence of disease. Likelihood ratios above 10 are considered to provide strong evidence for correct diagnoses in most circumstances (Koepsell & Connell 1985; Deeks & Altman 2004).

Figure 4.3. Distribution of a risk factor among cases and controls of disease to show calculation of likelihood ratios. The likelihood ratio of a test result represented by a point on the horizontal line is the height of the right hand curve (cases) divided by the height of the left hand curve (controls) at that point.



Receiver operating characteristic (ROC) curves provide a graphical approach to determine the optimal threshold value for a test with a range of possible values that maximises sensitivity and specificity. The sensitivity or detection rate is plotted against the false-positive (1-specificity) rate for each possible threshold value, thus generating a curve (see Figure 4.4). A good screening test combines a high detection rate with a low false-positive rate for a series of cut-off points. The area under the ROC curve (AUC) can be calculated and is equivalent to the probability that an individual with the disease selected at random has a higher value of the measure than an individual without the disease selected at random. An AUC close to 1 denotes a good diagnostic test, which is both sensitive and specific, as there is a high detection rate and a high specificity rate. An AUC value of 0.5 denotes a worthless predictive test, as there is a false-positive for every true-positive individual identified (Wald *et al.* 1994; Pepe *et al.* 2004).

Figure 4.4. Receiver operating characteristic curve for series of threshold values showing curves for a good screening test, a poor screening test and a typical continuous risk factor.



Since the majority of prior studies claiming a role for CRP in the prediction of coronary disease did not evaluate its performance in these terms, CRP was measured in stored samples from the prospective second Northwick Park Heart Study (NPHSII), in order to evaluate its potential utility in the prediction of CHD in middle aged men. The performance of CRP as a screening test was compared with other biomarkers and risk factors and with the Framingham risk score. The incremental value of adding CRP to the Framingham risk score was also evaluated. Finally, the results were compared with findings from prior studies and reasons for the limited performance of CRP as a screening test for CHD were explored.

4.3 Methods

4.3.1 Study subjects

The Northwick Park Heart Study II (NPHSII) cohort is a large prospective study of 3012 healthy Caucasian middle-aged (50- to 64-year-old) men, originally recruited in 1986. Nine general practices participated in the study, and all patients were free of a history of unstable angina, MI or evidence of silent infarction, coronary surgery, aspirin or anticoagulant therapy, cardiovascular disease, or malignant disease at the time of recruitment. Baseline

characteristics and demographic information were ascertained by means of a questionnaire completed at the beginning of the study. Every year patients who were still alive were recalled for interview so that coronary events could be identified, and lipid measurements were repeated. A routine ECG was done at the sixth examination. The endpoints were fatal coronary heart disease (CHD) events and non-fatal MI, coronary artery surgery and silent MI on follow-up ECG. CHD events were defined as sudden cardiac death or symptomatic MI (based on history, electrocardiography, cardiac enzymes, and pathology; classified by World Health Organisation criteria.

4.3.2 Measurement of CRP

Plasma CRP concentrations were measured using commercial high-sensitivity assays (R&D Systems or Kordia Life Sciences). Inter-assay and intra-assay coefficients of variation were <6.2% and <1.9% respectively, with a detection limit of 0.1 mg/L. Of 2,738 men with CRP measures, 1,541 were from samples obtained at baseline, 724 from samples obtained at year 1, 341 at year 2, 100 at year 3, 27 at year 4, and 5 at year 5. All analyses were adjusted for the year of measurement.

4.3.3 Statistical analyses

These analyses were mainly conducted by Ms. Jackie Cooper. I contributed to the preparation of tabular data and helped to conduct some of the analyses.

Statistical analysis was performed using STATA (Intercooled STATA Version 8.2, STATA Corporation, Texas). Log-transformations were conducted for data not normally distributed and for these variables, geometric means and approximate standard deviations were generated. The association of CRP with other risk factors was assessed by Pearson correlation using measurements made in the same year as the CRP measure where possible. Survival analysis was carried out using Cox proportional hazards models, with events occurring before the CRP measurement being excluded from the analysis, and the time of the CRP measurement being taken as the start of follow-up. To allow for differences according to age and practice, age was included as a covariate in the model and the data stratified by practice (using the strata option in STATA). In additional models, adjustment was made for established coronary risk factors as well as fibrinogen.

Means and standard deviations of log-transformed CRP, as well as certain other risk factors and biomarkers in those with and without CHD were used to construct normal curves to determine the overlap between the distributions in the two groups. Comparisons were then made with prior prospective studies evaluating the CRP-CHD association. Studies were chosen if they were prospective studies, with geometric mean and SD values for CRP concentration among

cases and controls. Distributions of CRP values in cases and controls were then reconstructed using these values. These were then used to estimate the disease detection rate, for a 5% and 10% false positive rate, using the method of Wald et al. (Wald *et al.* 2005). The performance of CRP as a screening test for CHD was assessed by means of disease detection and false positive rates, and by using the area under the receiver-operating characteristic (ROC) curve as a discriminatory test, with use of Harrell's C-index to allow for right-censored data (Pencina & D'Agostino 2004). The performance of the Framingham risk score (Wilson *et al.* 1998), a Kaplan-Meier model generated using Framingham variables, and the same model with the addition of CRP was evaluated by ranking participants according to fifths of predicted risk, and comparing predicted and observed 10-year risks of CHD within each category.

4.4 Results

4.4.1 Distribution of CRP concentration in the NPHSII cohort

Of the 3,012 men initially recruited, 2,479 had CRP measures available. The distribution of CRP was right skewed, with most subjects having a CRP concentration of less than 4mg/L (see Figure 4.5). The median CRP was 2.55mg/L (interquartile range 1.25-5.21mg/L). Log-transformation resulted in normalisation of the distribution with geometric mean 2.46mg/L (approximate SD 2.51mg/L). This distribution is similar to that seen in other observational studies conducted in different populations (see Figure 4.6).

Figure 4.5. CRP distribution in the NPHSII cohort taken from men where measures were available.

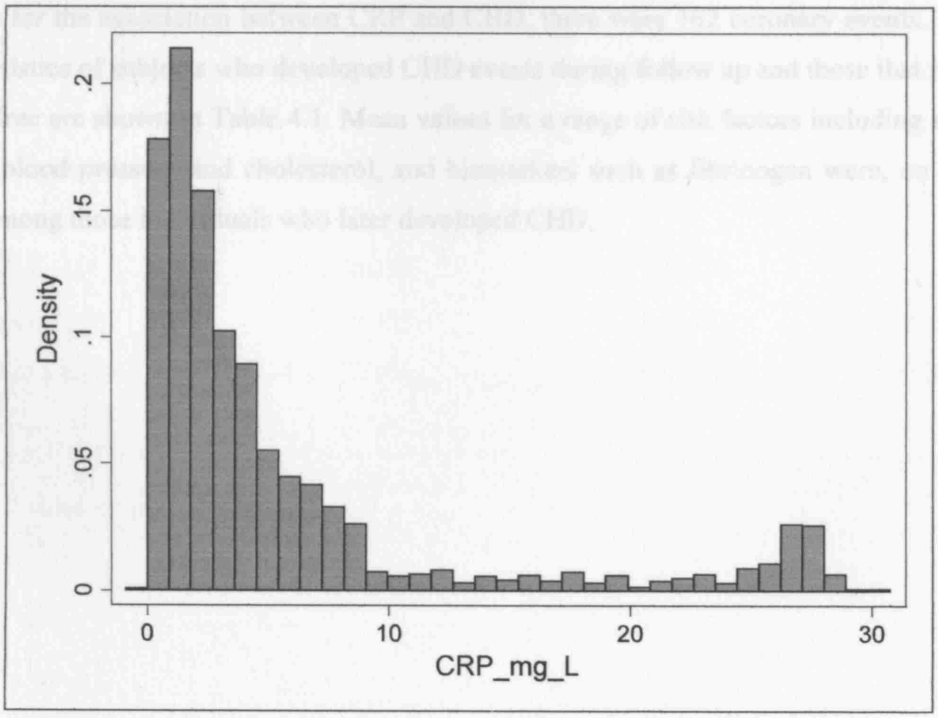
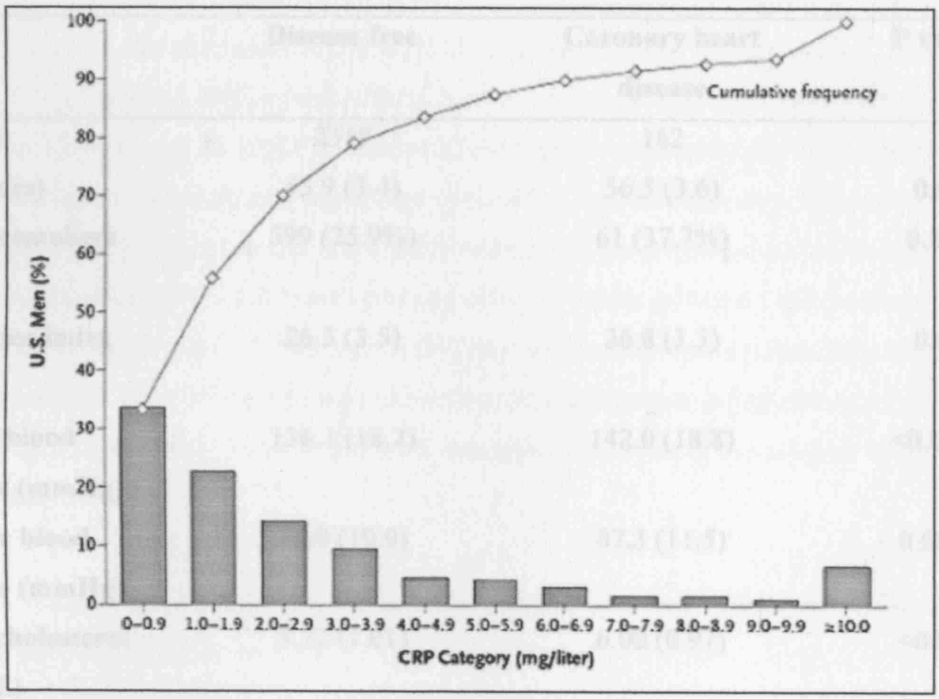


Figure 4.6. CRP distribution among 4969 men in the United States from the National Health and Nutrition Examination Survey (NHANES) (Woloshin & Schwartz 2005).



4.4.2 Baseline characteristics in subjects with later CHD and those that remained disease free

During an average follow-up of 10.8 years (9 days to 16 years), 227 coronary events were recorded (defined using World Health Organisation criteria). Among the 2479 men assessed for the association between CRP and CHD, there were 162 coronary events. Baseline characteristics of subjects who developed CHD events during follow up and those that remained disease free are shown in Table 4.1. Mean values for a range of risk factors including smoking, systolic blood pressure and cholesterol, and biomarkers such as fibrinogen were, on average, higher among those individuals who later developed CHD.

Table 4.1. Baseline characteristics of subjects who developed coronary heart disease and those who remained disease free.

	Disease free	Coronary heart disease	P value
Number	2317	162	
Age (years)	55.9 (3.4)	56.5 (3.6)	0.03
Current smokers (no./%)	599 (25.9%)	61 (37.7%)	0.001
Body mass index (kg/m^2)	26.3 (3.5)	26.8 (3.3)	0.07
Systolic blood pressure (mmHg)*	136.1 (18.2)	142.0 (18.8)	<0.0001
Diastolic blood pressure (mmHg)	84.0 (10.9)	87.3 (11.5)	0.0003
Serum cholesterol (mmol/L)	5.72 (1.01)	6.00 (0.97)	<0.001
HDL-cholesterol (mmol/L)	0.81 (0.25)	0.75 (0.21)	0.06
Triglycerides* (mmol/L)	1.75 (0.92)	2.11 (1.15)	<0.0001
Framingham risk (10 year risk of CHD)	Score=12.8 (2.0) Risk=12%	Score=13.7 (2.1) Risk=16%	<0.0001
Apolipoprotein B* (mg/dL)	0.85 (0.24)	0.91 (0.21)	0.002
Apolipoprotein A1 (mg/dL)	1.65 (0.32)	1.61 (0.26)	0.16
ApoB/A1 ratio*	0.52 (0.20)	0.57 (0.17)	0.007
Fibrinogen* (g/l)	2.66 (0.48)	2.83 (0.46)	<0.0001
C-reactive protein* (mg/L)	2.40 (2.44)	3.66 (3.66)	0.0004

*geometric mean (approx SD)

4.4.3 CRP and risk of coronary disease

The risk of CHD by tertile of CRP is shown in Table 4.2. The hazard ratios associated with increasing tertiles of CRP were 1.54 (95%CI: 1.02-2.34) and 2.61 (95%CI: 1.78-3.82) after age and practice-adjusted analyses ($p<0.0001$). However, risk was attenuated further after adjustment for smoking, BMI, systolic blood pressure, total and HDL cholesterol and fibrinogen (tertile 2 hazard ratio=1.23 (95%CI: 0.65-2.32) and tertile 3 hazard ratio=2.20 (95%CI: 1.18-4.11)), although it still remained significant ($p<0.02$). These hazard ratios for similar levels of adjustment are very similar to the odds ratios reported in the meta-analysis of Danesh et al. of 22 observational studies of CRP and coronary disease, which included 7068 incident coronary events (Danesh *et al.* 2004). When the CRP-CHD relationship was modelled by regression, a unit increment in log CRP was associated with a similar increase in log risk across the whole range of CRP values with no evidence for a threshold, or curvature or variation in the slope of the relationship.

Table 4.2. Hazard ratios for coronary heart disease among individuals with CRP values in the middle and top thirds of the CRP distribution.

	Hazard ratio (95% confidence interval)		
	Model 1	**Model 2**	**Model 3**
Coronary heart disease			
Tertile 1	1.00 (reference)	1.00 (reference)	1.00 (reference)
Tertile 2	1.54 (1.02–2.34)	1.36 (0.89-2.08)	1.23 (0.65-2.32)
Tertile 3	2.61 (1.78-3.82)	2.10 (1.42-3.12)	2.20 (1.18-4.11)
p for trend	0.0001	0.0004	0.02

Model 1: adjusted for age and practice.

Model 2: adjusted for practice and Framingham risk score.

Model 3: adjusted for age, practice, BP, smoking, total- and HDL- cholesterol, BMI and fibrinogen.

4.4.4 Performance of CRP and other measures as a screening test for CHD

Mean values for CRP, as well as a range of risk factors and biomarkers were higher on average, among subjects who later developed CHD (see Table 4.1). However for all these measures, there was a substantial overlap of the distributions between subjects who later developed CHD and those who remained disease free (see Figures 4.7-4.11).

In setting threshold values for disease detection or screening, higher CRP cut-off points that had acceptable false positive rates allowed detection of only a minority of all cases of CHD

(see Table 4.3). Lower thresholds were associated with higher detection rates but unacceptably high rates of misclassification, for example, a CRP cut-off that led to an 88% detection rate also had an 82% false positive rate (see Table 4.3). Using the values for the detection rate and false positive rate for a range of CRP values, a ROC curve was plotted to show the performance of CRP in detecting CHD events (see Figure 4.12). Similar ROC curves were plotted for other continuous measures including blood pressure, cholesterol and fibrinogen. The areas under the ROC curve (AUC), which have been calculated to give an estimate of the probability of determining whether a patient develops CHD for different cut points, are given in Table 4.4 along with the hazard ratio. The AUCs for all risk factors or biomarkers evaluated were similar, with no single measure exhibiting particularly high ability to discriminate subjects who later developed CHD. The hazard ratios provided little information on predictive performance as adjudged by the wide variation in hazard ratios with little or no difference in AUCs.

Because in prospective studies, the performance of screening tests depends on the duration of follow-up, analyses were repeated using a modification of the area under the ROC curve for use with so-called right-censored data, where the duration of follow-up varies among individuals. Positive and negative predictive values were obtained for a follow-up of ten years (see Table 4.5), and were similar to those seen in Table 4.3. C index values (comparable to AUC values) using this method were also very similar to the AUC values seen in Table 4.4 (see Table 4.6).

Figure 4.7. CRP distribution in cases and controls

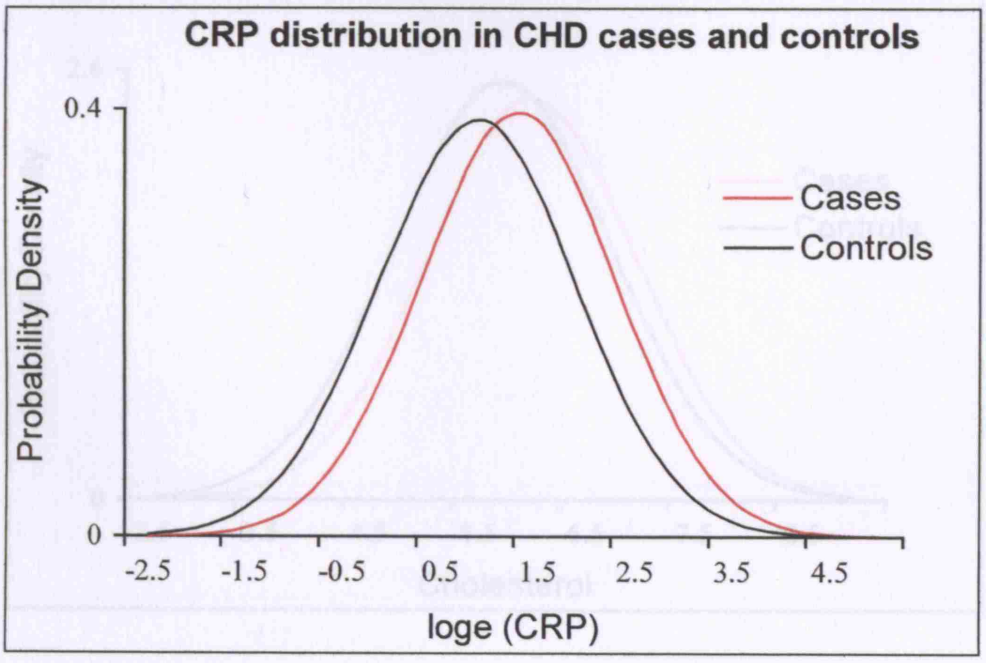


Figure 4.8. Systolic blood pressure distribution in cases and controls

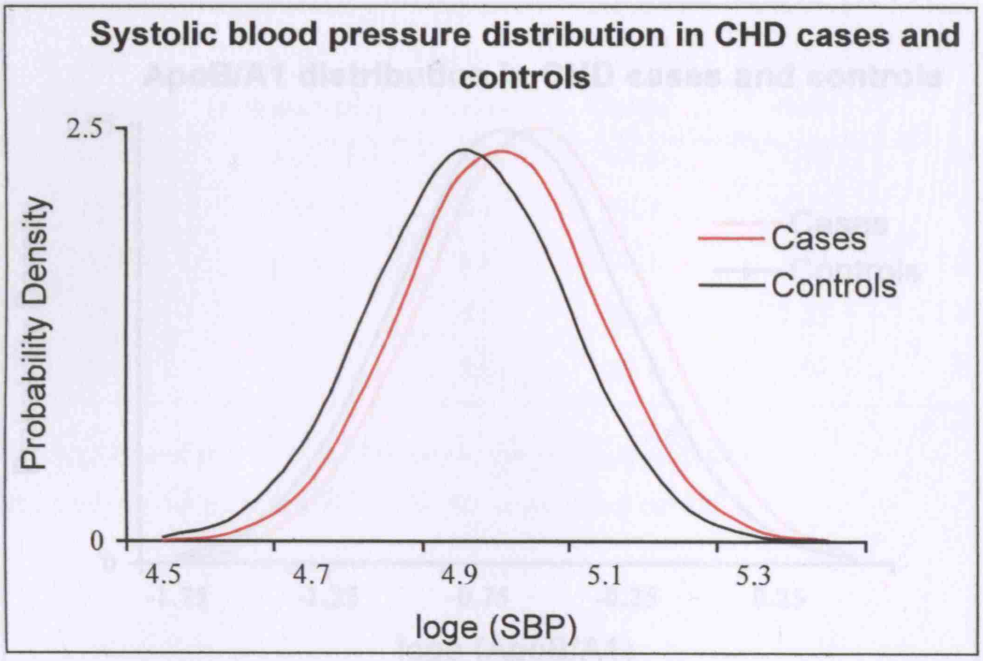


Figure 4.9. Cholesterol distribution in cases and controls

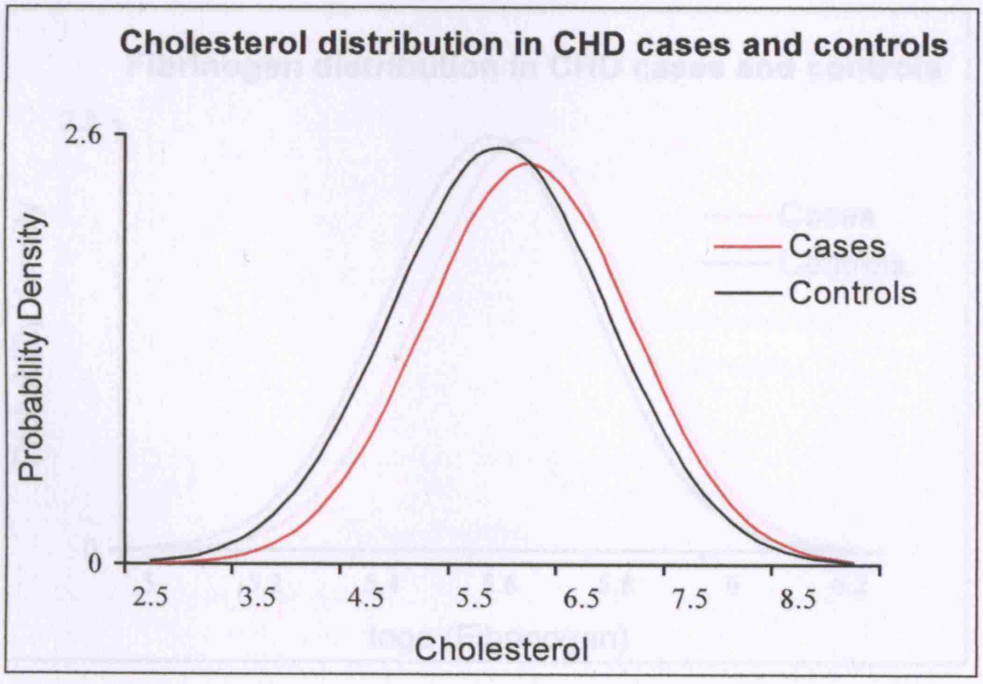


Figure 4.10. ApoB/ApoA1 distribution in cases and controls

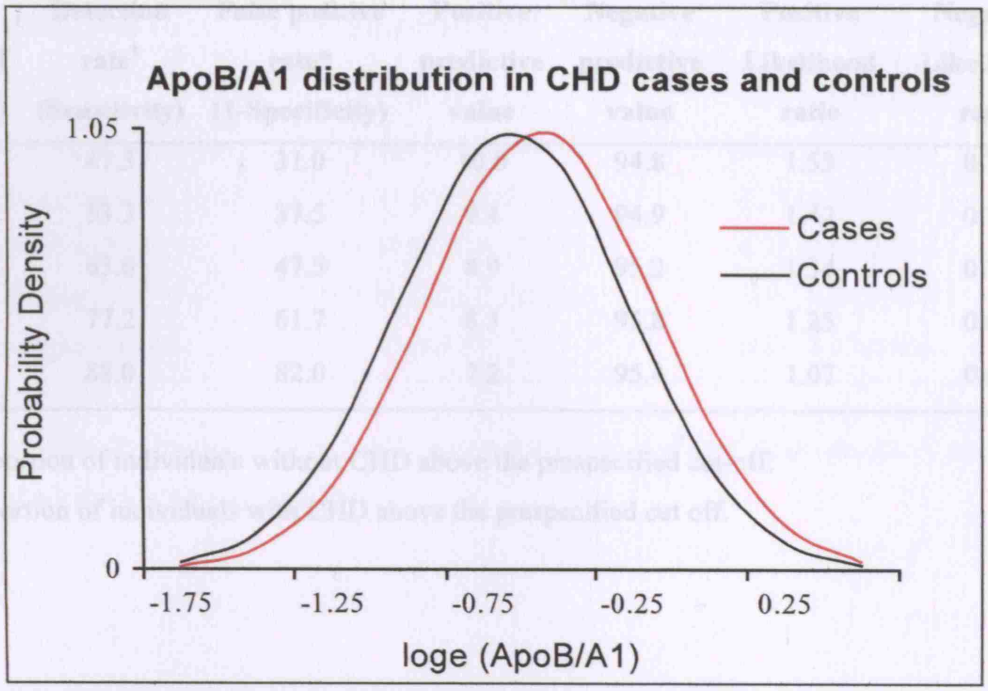


Figure 4.11. Fibrinogen distribution in cases and controls

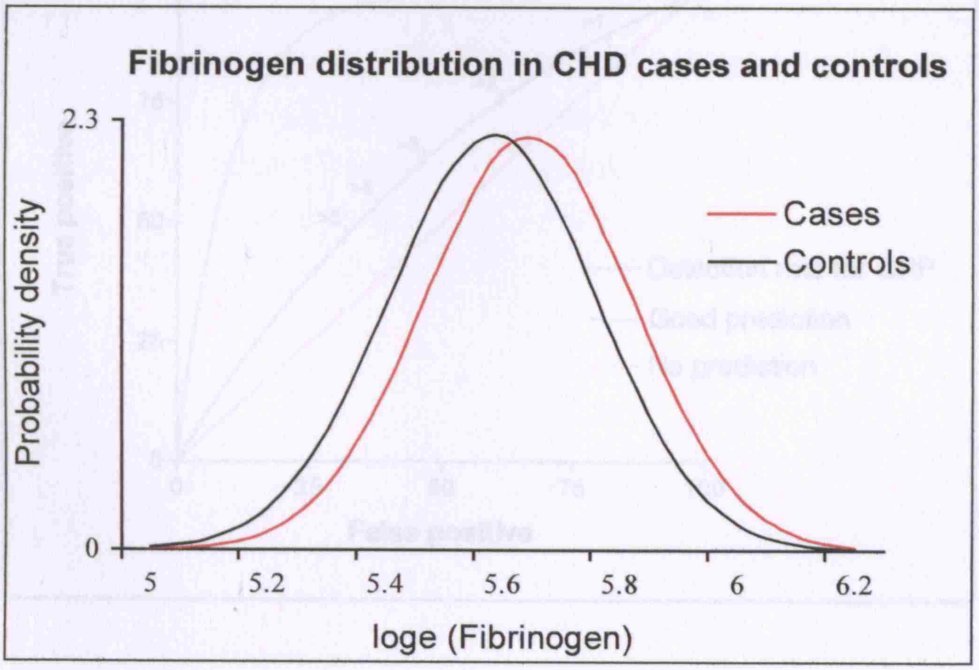


Table 4.3. Performance of CRP in detecting men who develop coronary heart disease.

CRP cut-off (mg/L)	Detection rate† (Sensitivity)	False positive rate* (1-Specificity)	Positive predictive value	Negative predictive value	Positive Likelihood ratio	Negative Likelihood ratio
> 5	47.3	31.0	10.0	94.8	1.53	0.76
> 4	53.3	37.5	9.4	94.9	1.42	0.74
> 3	63.6	47.5	8.9	95.2	1.34	0.70
> 2	77.2	61.7	8.3	95.8	1.25	0.60
> 1	88.0	82.0	7.2	95.4	1.07	0.67

*Proportion of individuals without CHD above the prespecified cut-off.

†Proportion of individuals with CHD above the prespecified cut off.

Figure 4.12. ROC curve to show performance of CRP in the detection of CHD.

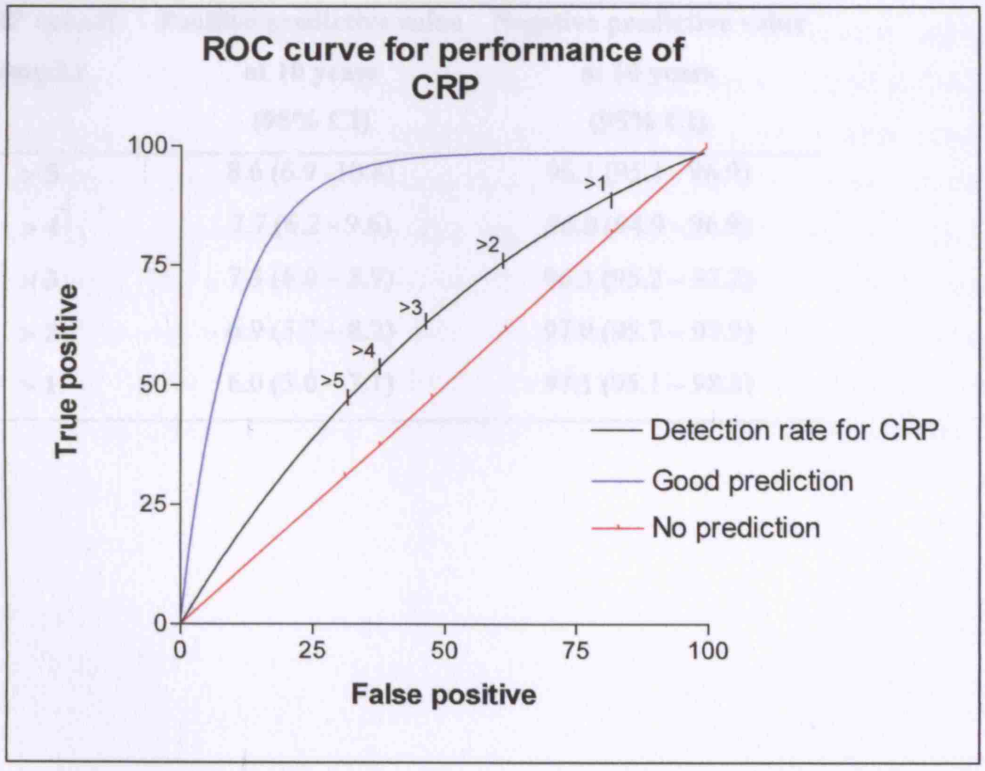


Table 4.4. Area under the ROC curve for various biomarkers and risk factors in isolation.

Screening test	Hazard ratio*	Area under ROC curve (95% confidence interval)
CRP	2.53 (1.44-4.62)	0.61 (0.57-0.66)
Systolic blood pressure	1.96 (1.10-3.50)	0.60 (0.55-0.64)
Total cholesterol	1.35 (0.75-2.43)	0.59 (0.55-0.64)
ApoB/A1 ratio	1.79 (0.77-4.17)	0.59 (0.54-0.63)
Fibrinogen	1.30 (0.74-2.28)	0.61 (0.56-0.65)
Framingham score	1.47 (0.68-3.18)	0.62 (0.57-0.66)

*Hazard ratio for top vs. bottom tertile adjusted for age, practice, blood pressure, total- and HDL- cholesterol where appropriate.

Table 4.5. Performance of CRP in detecting men who develop coronary heart disease using right-censored data.

CRP cut-off (mg/L)	Positive predictive value at 10 years (95% CI)	Negative predictive value at 10 years (95% CI)
> 5	8.6 (6.9 -10.8)	96.1 (95.1 - 96.9)
> 4	7.7 (6.2 - 9.6)	96.0 (94.9 - 96.9)
> 3	7.3 (6.0 – 8.9)	96.3 (95.2 – 97.2)
> 2	6.9 (5.7 – 8.2)	97.0 (95.7 – 97.9)
> 1	6.0 (5.0 – 7.1)	97.1 (95.1 – 98.3)

Table 4.6. C index to detect CHD events for various biomarkers and risk factors using right-censored data.

Screening test	C index (95% confidence interval)*
CRP	0.63 (0.60-0.66)
Systolic blood pressure	0.59 (0.56-0.62)
Total cholesterol	0.59 (0.56-0.62)
ApoB/A1 ratio	0.60 (0.57-0.63)
Fibrinogen	0.62 (0.59-0.64)
Framingham score	0.62 (0.60-0.65)
Framingham + CRP	0.66 (0.63-0.68)

*Predictive accuracy is calculated as Harrell's C-index which extends the area under the ROC curve to the case of right-censored survival data.

4.4.5 Distribution of CRP in NPHSII and published studies

Since there was a substantial overlap of the CRP distributions between subjects who later developed CHD (cases) and those who remained disease free (controls) in the NPHSII study, published data on CRP concentrations among cases and controls were examined to determine whether this overlap in distributions was specific to the NPHSII study or whether it was seen in all populations. Twenty-four prospective studies had data available on the geometric mean CRP concentrations and standard deviations in cases and controls (see Table 4.7). These data were used to infer the distribution in cases and controls on the assumption of normality (see Figures 4.13-4.16). In all studies examined, CRP distributions showed substantial overlap between cases and controls, which was concordant with the distributions seen in the NPHSII study.

Table 4.7. Published prospective studies of CRP.

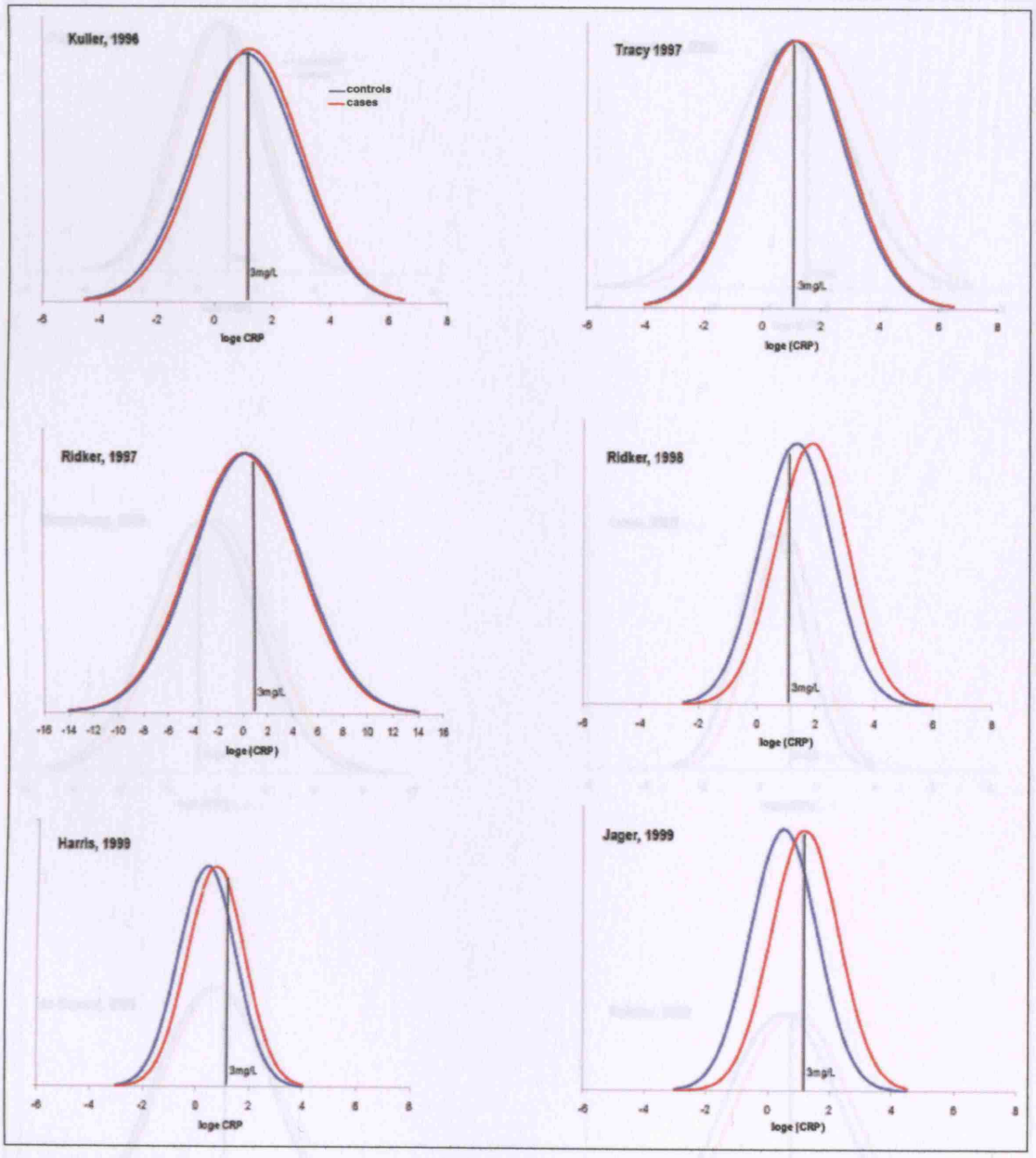
Study	Data set	% Men	Duration of follow-up (yrs)	CHD classification
Kuller, 1996	MRFIT	100	17	NA
Tracy, 1997	CHS	61	3	ECG/reported
Ridker, 1997	PHS	100	14	WHO criteria
Agewall, 1998	Goteborg	100	3	Fatal/non-fatal MI
Ridker, 1998	WHS	0	3	WHO criteria
Jager, 1999	Hoorn	48	5	WHO criteria
Koenig, 1999	MONICA	100	8	MONICA manual
Harris, 1999	RHS	41	4.6	ICD
Danesh, 2000	UK towns	100	12	WHO criteria
Mendall, 2000	CPHDS	100	13.7	WHO criteria
Packard, 2000	WOSCOPS	100	6	WHO criteria
Roivainen, 2000	HHS	100	8.5	ECG/reported
Strandberg, 2000	HAS	46	10	NYHA criteria
Lowe, 2001	Speedwell	100	6.25	WHO criteria
Pirro, 2001	QCS	100	5	ECG/reported
De Backer, 2002	Bellstress	100		ECG/reported
Folsom, 2002	ARIC	66	6	ECG/reported
Pradhan, 2002	WHI-OS	0	2.9	ECG/reported
Sakkinen, 2002	Honolulu	100	20	ECG/reported
Ridker, 2002	WHS	0	8	WHO criteria
Luc, 2003	PRIME	100	5	MONICA criteria
Van der Meer, 2003	Rotterdam	45.5	5	ICD
Danesh, 2004	Reykjavik	70.3	20.6	MONICA criteria
Koenig, 2004	MONICA	100	6.6	MONICA manual
Lowe, 2004	Speedwell	100	6.25	WHO criteria
Lowe, 2004	Caerphilly	100	8.75	WHO criteria
Pai, 2004	NHS/HPF	64	8	WHO criteria
Rutter, 2004	Fram Offs	45	7	WHO criteria
Bard, 2005	Michigan	75		NA
Boekholdt, 2005	EPIC-N	63	6	ICD
Cushman, 2005	CHS	42	10	ECG/reported
Lawlor, 2005	BWHHS	0	3.5	WHO criteria
Shlipak, 2005	CvHS	42.5	8.6	Reported
Wilson, 2005	FHS	45.8	8	NA
St. Pierre, 2005	QCS	100	13	ECG/reported
Stork, 2006	Zoetermer	100	4	Plaque assessed
Folsom, 2006	ARIC		5	ECG/reported
May, 2006	BWHHS	0	5	WHO criteria

NA = not available.

ICD = International Classification of Diseases.

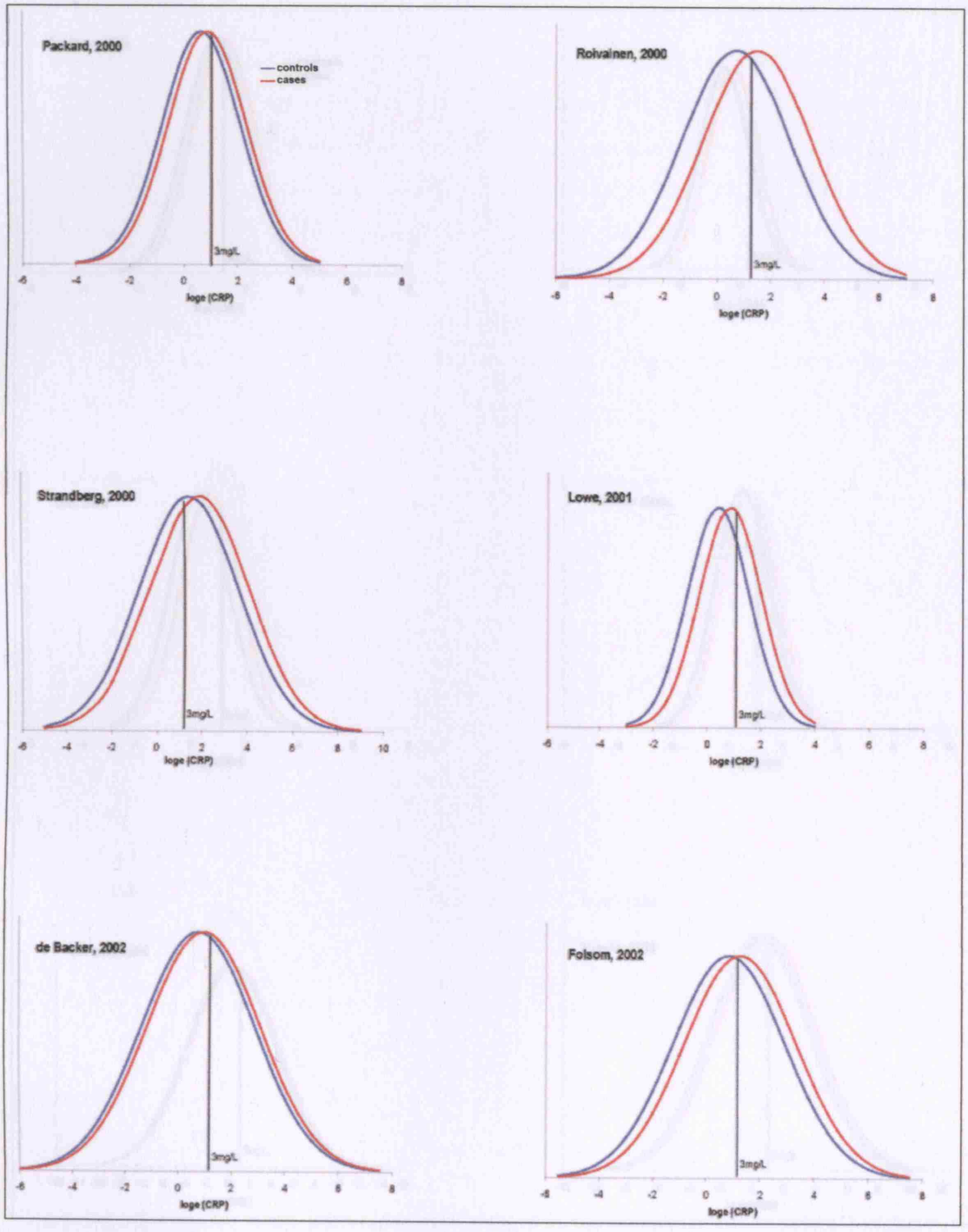
NYHA = New York Heart Association.

Figure 4.13. Inferred CRP distributions among cases and controls in studies from 1996 to 1999.



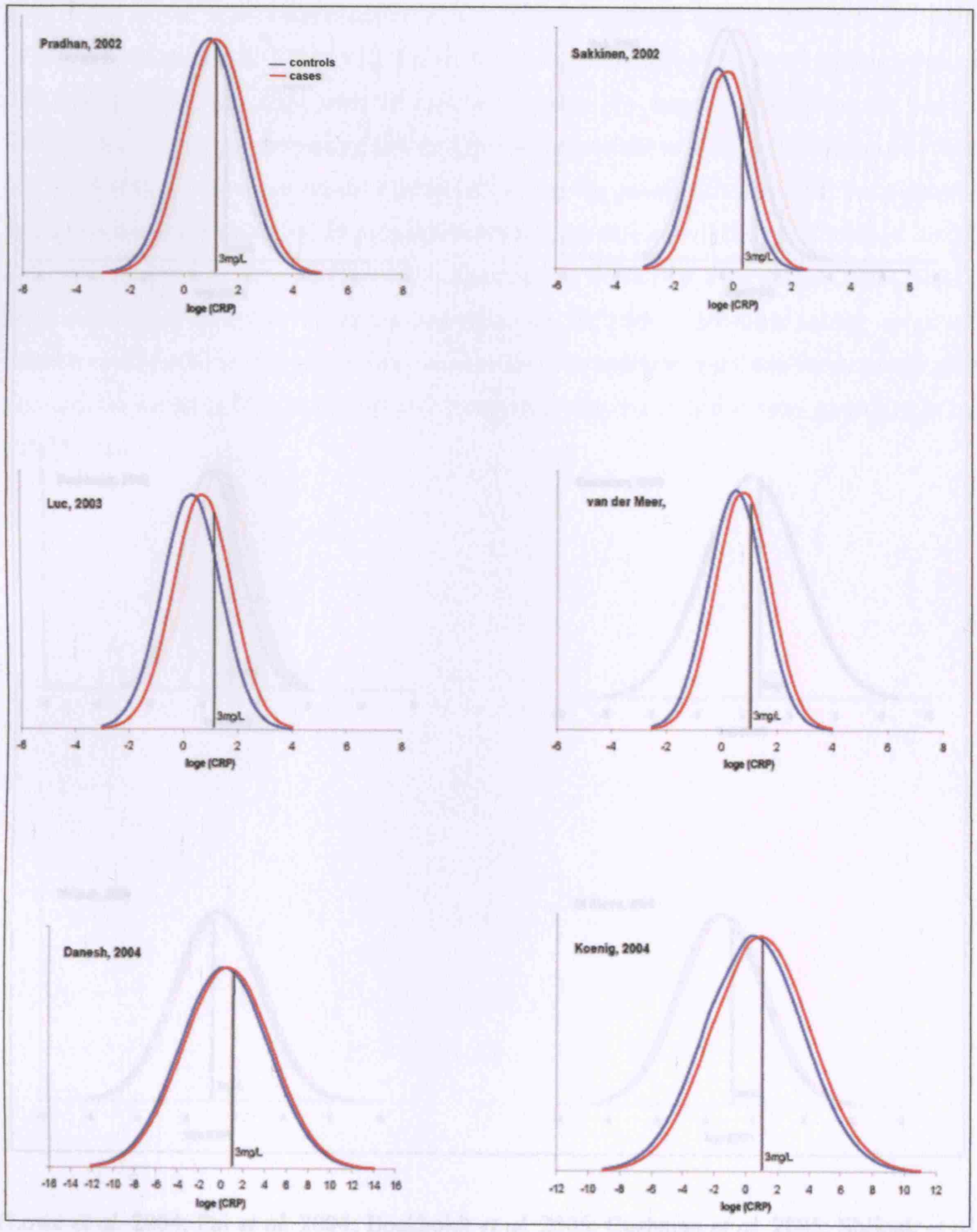
(Kuller *et al.* 1996; Tracy *et al.* 1997; Ridker *et al.* 1997; Ridker *et al.* 1998b; Harris *et al.* 1999; Jager *et al.* 1999).

Figure 4.14. Inferred CRP distributions among cases and controls in studies from 2000 to 2002.



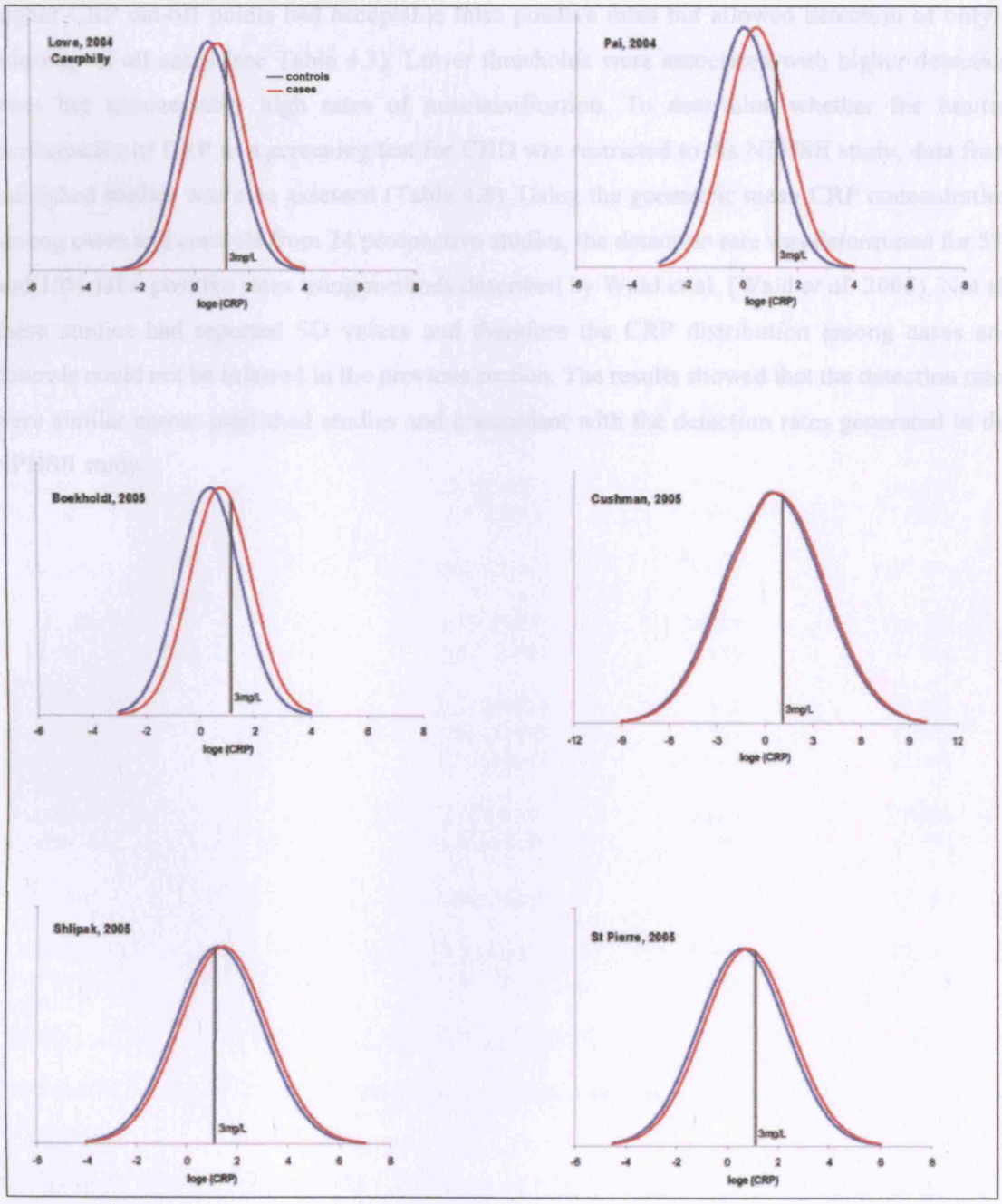
(Packard *et al.* 2000; Roivainen *et al.* 2000; Strandberg & Tilvis 2000; Lowe *et al.* 2001; De Backer *et al.* 2002; Folsom *et al.* 2002).

Figure 4.15. Inferred CRP distributions among cases and controls in studies from 2002 to 2004.



(Pradhan *et al.* 2002; Sakkinen *et al.* 2002; Luc *et al.* 2003; van der Meer *et al.* 2003; Danesh *et al.* 2004; Koenig *et al.* 2004).

Figure 4.16. Inferred CRP distributions among cases and controls in studies from 2004 to 2005.



(Lowe *et al.* 2004; Pai *et al.* 2004; Boekholdt *et al.* 2005; Cushman *et al.* 2005; Shlipak *et al.* 2005; St Pierre *et al.* 2005).

4.4.6 Performance of CRP as a screening test for CHD in NPHSII and published studies

In the NPHSII study, when setting threshold values for coronary disease detection, higher CRP cut-off points had acceptable false positive rates but allowed detection of only a minority of all cases (see Table 4.3). Lower thresholds were associated with higher detection rates but unacceptably high rates of misclassification. To determine whether the limited performance of CRP as a screening test for CHD was restricted to the NPHSII study, data from published studies was also assessed (Table 4.8). Using the geometric mean CRP concentration among cases and controls from 24 prospective studies, the detection rate was determined for 5% and 10% false positive rates using methods described by Wald et al. (Wald *et al.* 2005). Not all these studies had reported SD values and therefore the CRP distribution among cases and controls could not be inferred in the previous section. The results showed that the detection rates were similar across published studies and concordant with the detection rates generated in the NPHSII study.

Table 4.8. Performance of CRP in detecting CHD in NPHSII and published studies for 5% and 10% false positive rates.

Study	Mean CRP cases (mg/L) (N)	Mean CRP controls (mg/L) (N)	Detection rate (for 5% FPR)*	Detection rate (for 10% FPR)*
Kuller, 1996	3.4 (148)	2.9 (296)	6%	17.7%
Tracy, 1997	3.3 (107)	2.4 (107)	7%	13.3%
Ridker, 1997	1.37 (543)	1.10 (543)	5%	9.2%
Ridker, 1998	6.45† (122)	3.75† (244)	11.4%	20.1%
Jager, 1999	3.33† (58)	1.68† (573)	15.5%	25.8%
Harris, 1999	2.1 (323)	1.5 (353)	9.2%	16.7%
Packard, 2000	2.36 (580)	1.88 (1160)	6.8%	13.1%
Roivainen, 2000	4.44 (241)	2.01 (241)	10.5%	18.8%
Strandberg, 2000	6.55 (71)	4.05 (336)	7.7%	14.5%
Lowe, 2001	2.36 (162)	1.50 (1528)	10.6%	18.8%
De Backer, 2002	2.54 (446)	2.06 (892)	6.1%	11.9%
Folsom, 2002	3.5 (242)	2.3 (590)	7.5%	14.1%
Pradhan, 2002	3.3† (304)	2.5† (304)	7.9%	14.8%
Sakkinen, 2002	0.81 (369)	0.60 (1348)	9.8%	17.7%
Luc, 2003	2.0† (317)	1.33† (609)	10.2%	18.2%
Van der Meer, 2003	2.18† (157)	1.68† (500)	9.2%	17.3%
Danesh, 2004	1.75 (2406)	1.28 (3891)	5.8%	11.4%
Koenig, 2004	2.56 (191)	1.64 (3244)	6.7%	12.8%
Lowe, 2004 (Caerphilly)	2.26 (189)	1.66 (1334)	9%	16.4%
Pai, 2004	3.1† (239)	2.2† (469)	8.6%	15.8%
Boekholdt, 2005	2.1† (770)	1.4† (1515)	9.9%	17.9%
Cushman, 2005	1.98 (547)	1.64 (3424)	5.7%	11.1%
Shlipak, 2005	4.3 (1249)	3.5 (4559)	6.4%	12.3%
St. Pierre, 2005	2.2 (210)	1.8 (1772)	6.3%	12.2%
NPHSII	3.66 (162)	2.40 (2317)	10%	18%

†Median CRP values. *Detection rates were estimated using the method of Wald et al. (Wald *et al.* 2005).

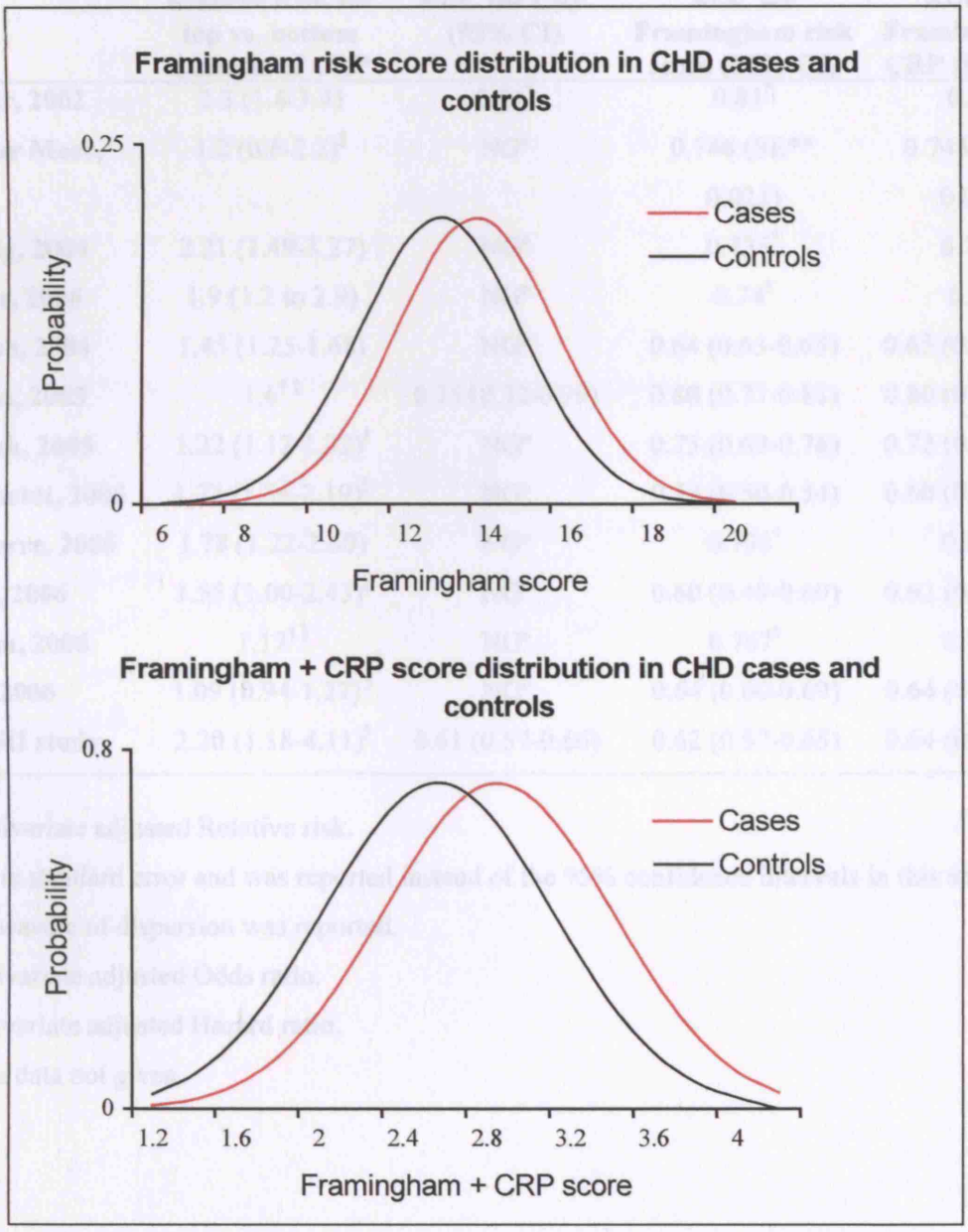
(Kuller *et al.* 1996; Tracy *et al.* 1997; Ridker *et al.* 1997; Ridker *et al.* 1998b; Harris *et al.* 1999; Jager *et al.* 1999); (Packard *et al.* 2000; Roivainen *et al.* 2000; Strandberg & Tilvis 2000; Lowe *et al.* 2001; De Backer *et al.* 2002; Folsom *et al.* 2002); (Pradhan *et al.* 2002; Sakkinen *et al.* 2002; Luc *et al.* 2003; van der Meer *et al.* 2003; Danesh *et al.* 2004; Koenig *et al.* 2004); (Lowe *et al.* 2004; Pai *et al.* 2004; Boekholdt *et al.* 2005; Cushman *et al.* 2005; Shlipak *et al.* 2005; St Pierre *et al.* 2005).

4.4.7 Combining CRP with the Framingham risk score

In clinical practice, the Framingham, PROCAM or other risk scores that utilise established risk factors are used widely to estimate CHD risk. These scores aim to improve

disease detection rates by combining information from several risk factors or markers rather than considering single risk factors in isolation. Prior studies have suggested that CRP provides useful additional predictive information at all levels of Framingham risk (Ridker *et al.* 2003; Ridker *et al.* 2003). It has been suggested that CRP estimation may have particular utility in prediction among individuals with an intermediate range of 10-year CHD risk of 10-20% (calculated using the Framingham risk equation (Wilson *et al.* 1998)). 72% of the men from the NPHSII data set had a Framingham risk score within this range. The distribution of Framingham risk scores among subjects who later did and did not develop CHD events was therefore examined in NPHSII and compared with the distribution when CRP was added to the risk score (see Figure 4.17).

Figure 4.17. Framingham risk score distribution in CHD cases and controls with, and without, CRP.



The effect of adding CRP to the Framingham risk score on screening performance was examined by evaluating the change in the AUC. There was a small change from 0.62 (0.57-0.66) to 0.64 (0.60-0.69) (p=0.10) after the addition of CRP. The effect of the addition of CRP to the Framingham risk score on the AUC from prior studies that reported these values are shown in Table 4.9, and are in line with the results from the current study (Ridker *et al.* 2002; van der Meer *et al.* 2003; Koenig *et al.* 2004; Rutter *et al.* 2004; Danesh *et al.* 2004; Wilson *et al.* 2005; Shlipak *et al.* 2005; Boekholdt *et al.* 2005; St Pierre *et al.* 2005; Stork *et al.* 2006; Folsom *et al.* 2006; May *et al.* 2006).

Table 4.9. Area under the ROC curve (AUC) values for the Framingham risk score distribution with and without CRP.

Study	Relative Risk for top vs. bottom quantile of CRP	AUC for CRP (95% CI)	AUC for Framingham risk score (95% CI)	AUC for Framingham + CRP (95% CI)
Ridker, 2002	2.3 (1.6-3.4)	0.64†	0.81†	0.81†
van der Meer, 2003	1.2 (0.6-2.2)‡	NG°	0.746 (SE** 0.021)	0.748 (SE** 0.021)
Koenig, 2004	2.21 (1.49-3.27)	NG°	0.735†	0.750†
Rutter, 2004	1.9 (1.2 to 2.9)	NG°	0.74†	0.74†
Danesh, 2004	1.45 (1.25-1.68)	NG°	0.64 (0.63-0.65)	0.65 (0.64-0.67)
Wilson, 2005	1.6†‡	0.75 (0.72-0.79)	0.80 (0.77-0.83)	0.80 (0.77-0.83)
Shlipak, 2005	1.22 (1.12-1.32)‖	NG°	0.73 (0.69-0.76)	0.72 (0.68-0.75)
Boekholdt, 2005	1.72 (1.36-2.19)‡	NG°	0.52 (0.50-0.54)	0.60 (0.58-0.62)
St. Pierre, 2005	1.78 (1.22-2.60)	NG°	0.705†	0.706†
Stork, 2006	1.55 (1.00-2.43)‖	NG°	0.60 (0.49-0.69)	0.62 (0.50-0.74)
Folsom, 2006	1.17†‖	NG°	0.767†	0.770†
May, 2006	1.09 (0.94-1.27)‖	NG°	0.64 (0.60-0.69)	0.64 (0.60-0.69)
NPHSII study	2.20 (1.18-4.11)‖	0.61 (0.57-0.66)	0.62 (0.57-0.65)	0.64 (0.60-0.69)

*Multivariate adjusted Relative risk.

**SE is standard error and was reported instead of the 95% confidence intervals in this study.

†No measure of dispersion was reported.

‡Multivariate adjusted Odds ratio.

‖Multivariate adjusted Hazard ratio.

°NG is data not given.

Individuals were then divided into fifths of predicted risk, using either the Framingham risk equation (Table 4.10), a Kaplan-Meier model derived using Framingham variables (Table 4.11), and the same model but with the addition of CRP (Table 4.12). As expected, a study-specific model derived using Framingham variables performed better than the Framingham equation itself, but the addition of CRP to this model produced no material improvement in the ability to categorise risk

Table 4.10. Calibration estimates for the Framingham risk equation.

Fifth of risk	Number of subjects	Number of events	% Predicted (at 10 years)	% Observed	Observed ten year risk (K-M)	Predicted/ Observed K-M
1	497	17	7.9	3.4	3.2	2.47
2	495	23	10.9	4.7	3.2	3.41
3	496	25	13.5	5.0	4.7	2.87
4	496	30	16.4	6.1	5.7	2.88
5	495	67	21.6	13.5	13.0	1.66
Total	2479	162	13.5	6.5	5.9	2.29

Table 4.11. Calibration estimates for model in NPHSII using Framingham variables.

Fifth of risk	Number of subjects	Number of events	% Predicted (at 10 years)	% Observed	Observed ten year risk (K-M)	Predicted/ Observed K-M
1	496	18	2.8	3.6	3.0	0.93
2	496	25	4.0	5.0	4.0	1
3	496	23	5.2	4.6	4.3	1.21
4	496	31	6.8	6.3	5.5	1.24
5	495	65	11.2	13.1	13.0	0.86
Total	2479	162	6.0	6.5	5.9	1.02

Table 4.12. Calibration estimates for model in NPHSII using Framingham variables and CRP.

Fifth of risk	Number of subjects	Number of events	% Predicted (at 10 years)	% Observed	Observed ten year risk (K-M)	Predicted/ Observed K-M
1	496	16	2.2	3.2	2.6	0.85
2	496	21	3.6	4.2	3.0	1.2
3	496	23	5.1	4.6	4.0	1.28
4	496	37	7.1	7.5	7.1	1
5	495	65	12.2	13.1	13.2	0.92
Total	2479	2479	6.0	6.5	5.9	1.02

4.4.8 Association of CRP with established and emerging risk factors

To explore the reasons for the poor incremental predictive performance of CRP when added to the Framingham risk score, the associations of CRP with components of the Framingham risk equation and other biomarkers relevant to cardiovascular disease were evaluated. Table 4.13 shows the values of a range of risk factors and biomarkers among individuals divided according to tertiles of the CRP distribution.

Table 4.13. Risk factors by tertile of CRP, using measurements made in same year as the CRP measurements where possible. Smoking, alcohol, ApoA and ApoB are baseline measures.

	Tertile of CRP			
	1	2	3	P value
	<1.64mg/L	1.64 – 3.99mg/L	>3.99mg/L	
	N=827	N=827	N=825	
Age	56.3 (3.5)	56.6 (3.6)	57.0 (3.6)	<0.001
BMI	25.1 (3.1)	26.7 (3.3)	27.3 (3.6)	<0.0001
Smoking	18.5% (153)	24.8% (205)	36.6% (302)	<0.0001
SBP*	130.6 (17.0)	135.3 (17.9)	137.6 (18.6)	<0.0001
Alcohol (Median units/wk [IQR])	6 [1-14]	7 [2-16]	7 [2-20]	0.01
Cholesterol	5.48 (0.91)	5.77 (1.00)	5.80 (0.99)	<0.0001
ApoA	1.70 (0.34)	1.65 (0.31)	1.60 (0.29)	<0.0001
ApoB*	0.81 (0.23)	0.86 (0.23)	0.89 (0.23)	<0.0001
ApoB/ApoA ratio*	0.48 (0.20)	0.53 (0.18)	0.57 (0.21)	<0.0001
HDL*	0.87 (0.26)	0.79 (0.24)	0.76 (0.23)	<0.0001
Triglyceride*	1.48 (0.73)	1.80 (0.91)	2.00 (1.07)	<0.0001
Framingham risk score	12.2 (2.0)	12.9 (2.0)	13.4 (2.0)	<0.0001
Fibrinogen*	2.46 (0.41)	2.63 (0.43)	2.87 (0.47)	<0.0001
Homocysteine*	12.0 (3.7)	12.1 (3.3)	12.5 (3.7)	0.16
Lp(a) (median & IQR)	8.1 [2.9 – 26.3]	7.8 [2.4-22.6]	10.0 [2.9-28.0]	0.0001

*Geometric mean (approx SD)

Pearson correlation coefficients between CRP and a subset of these measures are shown in Table 4.14. Thus, CRP exhibited correlation with many of the components of the Framingham risk score, as well as with other biomarkers of potential relevance to CHD risk.

Table 4.14. Correlations of risk factors and biomarkers with CRP.

	Correlation	P value
Age	r= 0.08	<0.0001
BMI*	r= 0.28	<0.0001
Systolic BP*	r= 0.18	<0.0001
Cholesterol	r= 0.16	<0.0001
Triglyceride*	r= 0.25	<0.0001
Fibrinogen*	r= 0.40	<0.0001

*Log-transformed.

Fibrinogen is adjusted for age and smoking.

4.5 Discussion

Many epidemiological studies have shown an association between CRP concentrations and risk of future cardiovascular morbidity among individuals at high risk, or among those with documented vascular disease (Pearson *et al.* 2003a; Muir *et al.* 1999; Heeschen *et al.* 2000). In addition to this, a number of prospective studies have demonstrated that even mild elevations of CRP concentrations among apparently healthy individuals are associated with higher risk for future cardiovascular events, leading to the proposal that CRP might provide an adjunct to coronary risk assessment (Ridker & Haughie 1998; Ridker *et al.* 2002; Danesh & Pepys 2000; Danesh *et al.* 1998).

In the current analysis in around 2500 men from the NPHSII study, the distribution of CRP, its association with cardiovascular risk factors and biomarkers, and its association with later coronary events was closely similar to that observed in prior studies and meta-analyses. However, when the predictive utility of CRP as a screening test for CHD was evaluated using the appropriate measures, its performance was limited, though somewhat similar to blood pressure and cholesterol.

The key issue for a predictive test is how well it distinguishes individuals who will or will not develop clinical events. It is interesting to note that of 37 prior studies (up to July 2006) that concluded that measurement of CRP could be useful in coronary risk prediction, only 12 studies have assessed its screening performance using the appropriate measures (Ridker *et al.*

2002; van der Meer *et al.* 2003; Koenig *et al.* 2004; Rutter *et al.* 2004; Wilson *et al.* 2005; Shlipak *et al.* 2005; Boekholdt *et al.* 2005; St Pierre *et al.* 2005; Stork *et al.* 2006; Folsom *et al.* 2006; May *et al.* 2006). Most others drew their conclusions based on relative measures of risk, such as odds ratios or hazard ratios, which while providing a useful measure of the strength of association, are not the appropriate metric for assessing the performance of a screening test. In the NPHSII cohort, hazard ratios for incident CHD, by tertile of CRP, were similar to the odds ratios reported in a recent meta-analysis of 22 observational studies involving 7068 incident cases (Danesh *et al.* 2004). Despite this, when the predictive utility of CRP was assessed using detection rates, false positive rates and ROC curves, CRP appeared to be a poor discriminator of subjects who do or do not develop coronary events.

The reasons for the limited performance of CRP as a screening test are likely to rest on two issues. A powerful screening test should have a threshold value that combines a high detection rate with a low misclassification (false positive) rate. The relationship between CRP and disease risk is graded and continuous, with no threshold value, below which risk is minimal and above which it is substantial. In addition, although the distribution of CRP is skewed, CRP values when log transformed are distributed approximately normally, with the majority of subjects in any population having intermediate values of CRP and hence an intermediate risk of coronary events. Therefore, most incident cases appear to occur among subjects without markedly elevated values of CRP, resulting in a substantial overlap in the distributions of cases and controls. Due to this overlap, it was not possible to find a cut-off value with a reasonable detection rate but low misclassification rate. Raising the threshold value reduced misclassification but led to extremely poor detection rates. When detection rates from prior studies were inferred by reconstructing distributions of CRP among cases and controls using the geometric means and approximate SD values reported in the primary studies, very similar findings were obtained (Kuller *et al.* 1996; Tracy *et al.* 1997; Ridker *et al.* 1997; Ridker *et al.* 1998b; Harris *et al.* 1999; Jager *et al.* 1999); (Packard *et al.* 2000; Roivainen *et al.* 2000; Strandberg & Tilvis 2000; Lowe *et al.* 2001; De Backer *et al.* 2002; Folsom *et al.* 2002); (Pradhan *et al.* 2002; Sakkinen *et al.* 2002; Luc *et al.* 2003; van der Meer *et al.* 2003; Danesh *et al.* 2004; Koenig *et al.* 2004); (Lowe *et al.* 2004; Pai *et al.* 2004; Boekholdt *et al.* 2005; Cushman *et al.* 2005; Shlipak *et al.* 2005; St Pierre *et al.* 2005).

In current clinical practice, risk scores that combine several risk factors are used to identify individuals at high coronary disease risk, in effect, to produce an ordered queue of subjects to help guide treatment, which tends to be targeted to individuals at highest risk, as this approach leads to fewer individuals treated per coronary event prevented. This approach was cost-effective when drugs like statins were expensive, but with the recent patent expiry of simvastatin, a more widespread, population-based approach to coronary prevention is being

debated (Wald & Law 2003). The Framingham risk equation is among the most widely used of these risk tools (Wilson *et al.* 1998). However, any risk score developed in one population may perform less well in another population (Brindle *et al.* 2003). The performance can be evaluated by comparing the observed and expected event rates among groups of individuals stratified by category of risk. When the observed vs. expected event rates were compared among individuals in NPHSII stratified by fifths of baseline risk, models utilising Framingham variables performed better than the Framingham equation itself. This is because the accuracy of the Framingham equation varies among studies as a result of varying background risk of the population under study and differences in the precision and reliability with which measures of cholesterol, blood pressure, smoking etc. were made. A recent systematic review examining the accuracy of the Framingham equation found a tendency to over predict absolute risk in populations with low observed CHD mortality and to under predict risk in populations with high CHD mortality (Brindle *et al.* 2006). When the utility of the Framingham risk score as a screening test was examined, its performance was also limited and similar to that seen for single risk factors. Therefore, even risk scores that combine information on multiple risk factors may be only moderately discriminatory. This is endorsed by the work in this chapter, which shows that many cases of CHD occur among individuals with average Framingham risk.

It has been suggested that CRP might provide additional predictive value when added to the Framingham risk score, particularly for subjects at intermediate levels of Framingham risk (10 year CHD risk of 10-20%). 72% of subjects from the NPHSII study had Framingham risk scores within this range, yet CRP added little to the predictive performance as shown by the minimal improvement in the area under the ROC curve (Δ area under ROC=0.02, p=0.10). This difference was not statistically significant, and the absolute difference is too small to suggest that CRP is likely to have clinical use as a screening tool for CHD. The reason for this may be partly due to the degree of cross-correlation between CRP and other measures that contribute to the Framingham risk score. This study therefore strongly suggests that CRP is at best, a moderate predictor of CHD events and does not provide incremental predictive information when added to the Framingham risk score.

4.6 Conclusions

CRP is a major acute phase reactant whose plasma concentrations increase markedly in response to infective or inflammatory processes. This has led to its wide use in clinical practice as a marker of infection or inflammation. Numerous observational studies have shown that CRP is also associated with risk of future cardiovascular disease, leading to the proposal that CRP measurement may have clinical utility in risk stratification. The aim of this study was to confirm this association between CRP and CHD events in a cohort of initially healthy men. The

performance of CRP in the detection of later CHD events was then evaluated and compared with the performance of other risk factors and biomarkers. It was also determined whether CRP added useful predictive information over and above traditional risk factors and the Framingham risk score.

This study found that the distribution of CRP is similar to that seen in other observational studies. The association between CRP and CHD events was confirmed and was similar in magnitude to other observational studies. However, when the performance of CRP in the detection of coronary events was assessed using ROC curve analyses, it appeared to be a poor discriminator of subjects who do or do not develop coronary disease. When CRP was added to the Framingham risk score to see if it might provide additional predictive value, there was minimal improvement in the area under the ROC curve, showing CRP added little to the predictive performance. This study and a re-evaluation of the published literature suggest that CRP is likely to have limited utility in the prediction of coronary disease. Nevertheless, the association of CRP with later coronary events could reflect a causal role for CRP in the development of heart disease. Therefore, in the remainder of this thesis, further studies were undertaken linking a genetic and epidemiological approach to help understand the causal relevance of CRP in coronary heart disease, using the principle of Mendelian randomisation.

5. Development of a consensus SNP and haplotype map of the human CRP gene

5.1 Aims

To construct a consensus genetic map of variants within and flanking the CRP gene. To examine linkage disequilibrium within the CRP gene to facilitate the generation of tagging SNPs for association studies. To assess whether promoter variants are likely to have functional effects by virtue of their location in consensus sites for transcription-factor binding.

5.2 Background

Information on common genetic variants in and around genes of interest can come from a variety of sources. Results of polymorphism screening or resequencing of candidate genes are sometimes reported in routine scientific publications and these were the main source of information on CRP gene variants until around 2002. Since then, a number of public domain resources have reported information on common variants across the genome as a whole, or across candidate genes of interest for a particular disease area in a more systematic manner. This has been complemented by information from the genome project and its spin-offs including the human HapMap.

5.2.1 Public domain databases with sequence information on the CRP gene

Public domain databases can vary in the amount of information available on a particular gene. Some databases provide information on SNPs across the genome, and usually result in the production of low-density SNP maps of hundreds of genes throughout the genome, such as those seen in the NCBI Single Nucleotide Polymorphism database (dbSNP), or the HapMap Consortium database and facilitate genome-wide association studies. Other databases, such as the University of Washington and the Fred Hutchinson Cancer Research Center (UW-FHCRC) Variation Discovery Resource (SeattleSNPs) and related Programs for Genomic Applications (PGA) sites use a fine-mapping technique for re-sequencing of specific candidate genes, and are thus able to produce high-density SNP maps of genes of interest for specific disease such as hypertension, obesity and CHD, or specific mechanistic pathways such as inflammation and clotting.

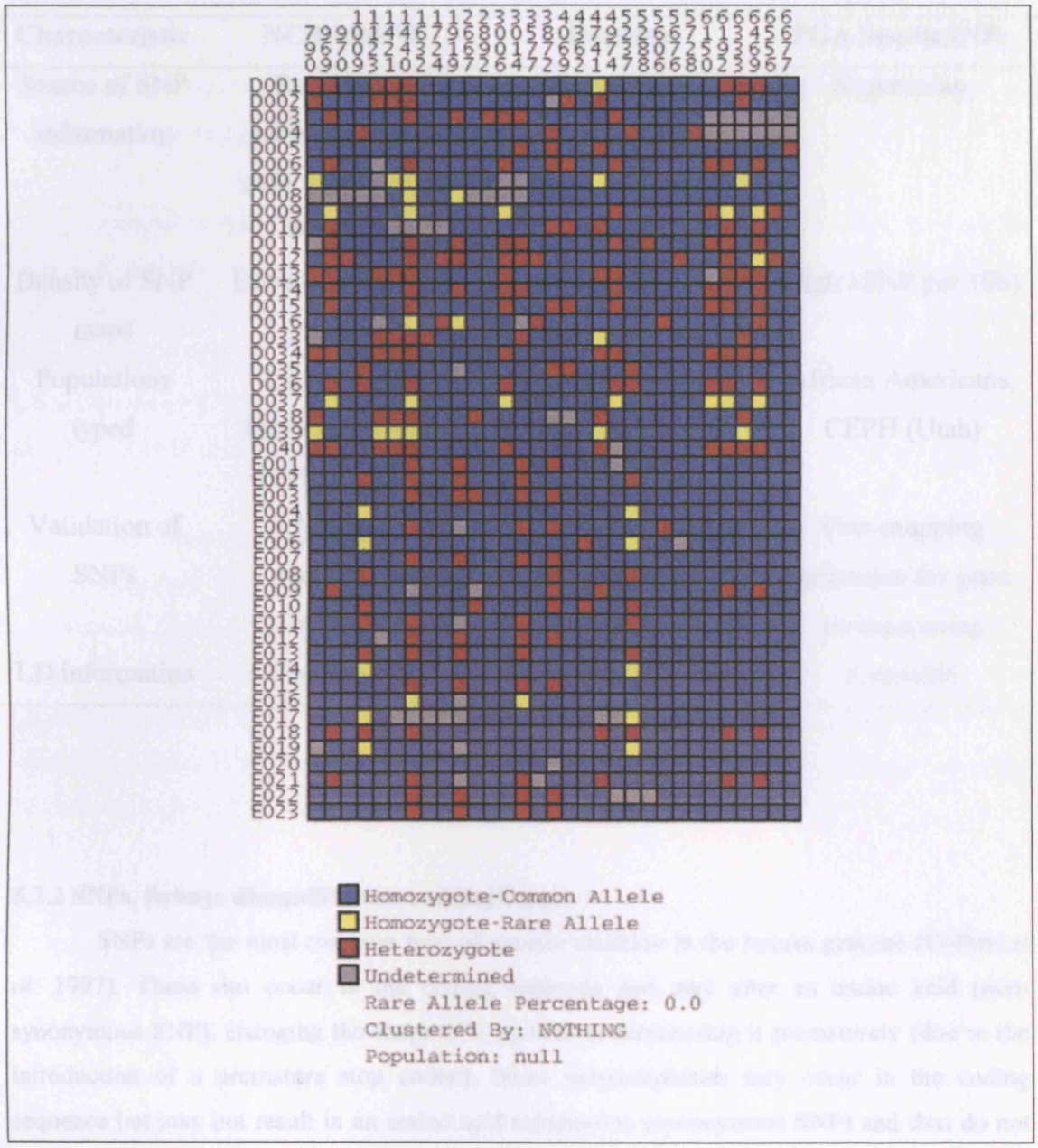
The NCBI Single Nucleotide Polymorphism database (dbSNP) is a collaboration between the National Human Genome Research Institute and the National Center for Biotechnology Information, and serves as a public-domain archive for a broad collection of simple genetic polymorphisms. This collection of polymorphisms includes single-base nucleotide substitutions (SNPs), small-scale multi-base deletions or insertions (DIPs), and microsatellite repeat variations (STRs). Each dbSNP entry includes the sequence context of the

polymorphism, the frequency of the polymorphism, and the experimental methods used to assay the variation. Polymorphisms are considered to be validated when confirmed either by multiple, independent submissions to the reference SNP cluster, by frequency or genotype data where minor alleles have been observed in at least two chromosomes, or by submitter confirmation. The NCBI SNP database also gives information about each SNP (where possible), such as the Hardy-Weinberg probability, the average estimated heterozygosity and the average allele frequency. These have been calculated by genotyping families from the Centre d'Etude du Polymorphisme Humain (CEPH), consisting of 30 mother-father-child trios that are Utah residents with ancestry from northern and western Europe, the HapMap collaboration, and the NIH Polymorphism Discovery Resource (NIHPDR) that consists of DNA from 450 anonymous, unrelated individuals with equal numbers of females and males.

The HapMap database is a culmination of information gathered from genotyping over one million SNPs across the human genome and resulting in the production of low-density SNP maps of hundreds of genes across the genome. A total of four geographical populations have been genotyped in 270 individuals. These are 30 trios (both parents and one child) of Yoruba people from Ibadan, Nigeria (YRI), 45 unrelated Japanese people from Tokyo (JPT), 45 unrelated Chinese from Beijing (CHB), and 30 CEPH trios (both parents and one child) (CEU). The HapMap database gives information about each SNP, including the Hardy-Weinberg probability, the average allele frequency and the pairwise LD between SNPs.

The Programs for Genomic Applications (PGAs) database was set up in 2000 by the National Heart, Lung, and Blood Institute (NHLBI) with the mission of advancing functional genomic research related to heart, lung, blood, and sleep disorders. The University of Washington and the Fred Hutchinson Cancer Research Center (UW-FHCRC) Variation Discovery Resource (SeattleSNPs) collaboration is one of the eleven PGAs, which differ according to disease interest, pathway interests and genes specific to these interests. The SeattleSNPs database has identified common variable sites in genes relevant to these disorders, and established their relative frequencies and haplotypes in two human populations with different evolutionary histories (African and European). The African descent population consists of 24 individuals (12 male/12 female) and is composed of DNA available from the Coriell Cell Repository (http://locus.umdnj.edu/). These individuals were selected from a human variation panel of 50 African Americans. The European descent population consists of 23 individuals (12 male/11 female) and is composed of CEPH Parent DNA, also available from the Coriell Cell Repository. The database provides a visual genotype for all the SNPs for each individual (see Figure 5.1), and uses this data to obtain the allele frequency and the Hardy-Weinberg probability.

Figure 5.1. Visual genotype in the CRP gene for individuals from the SeattleSNPs database (http://pga.gs.washington.edu/data/crp/crprt.prettybase.png).



Individuals are labelled from D001 to E023; polymorphisms are labelled from 790 to 6677.

Table 5.1. Characteristics of the NCBI dbSNP, HapMap and PGA SeattleSNPs databases of common variation.

Characteristic	NCBI dbSNP	HapMap	PGA SeattleSNPs
Source of SNP information	Sequencing, submission from other laboratories, computed	Sequencing	Sequencing
Density of SNP maps	Intermediate (1SNP per 600b)	Low (1SNP per 5kb)	High 1SNP per 10b)
Populations typed	CEPH (Utah), HapMap subjects, NIHPDR panel	Nigerian, Japanese, Chinese, CEPH (Utah)	African Americans, CEPH (Utah)
Validation of SNPs	Multiple, independent submissions	Sequencing of several subjects	Fine-mapping technique for gene re-sequencing
LD information	Not available	Available	Available

5.2.2 SNPs, linkage disequilibrium and haplotypes

SNPs are the most common type of genetic variation in the human genome (Collins *et al.* 1997). These can occur in the coding sequence and may alter an amino acid (non-synonymous SNP), changing the shape of a protein, or terminating it prematurely (due to the introduction of a premature stop codon). Some polymorphisms may occur in the coding sequence but may not result in an amino acid substitution (synonymous SNP) and thus do not change the composition of the protein. Polymorphisms that occur within the promoter region of a gene may disrupt transcription-factor binding sites required for efficient mRNA production, resulting in decreased mRNA transcription and reduced protein production without altering the structure and function of the protein product. Although most SNPs have no observable phenotype, they occur sufficiently frequently within the genome to act as markers that are useful for the identification of disease genes and genes underlying trait variation.

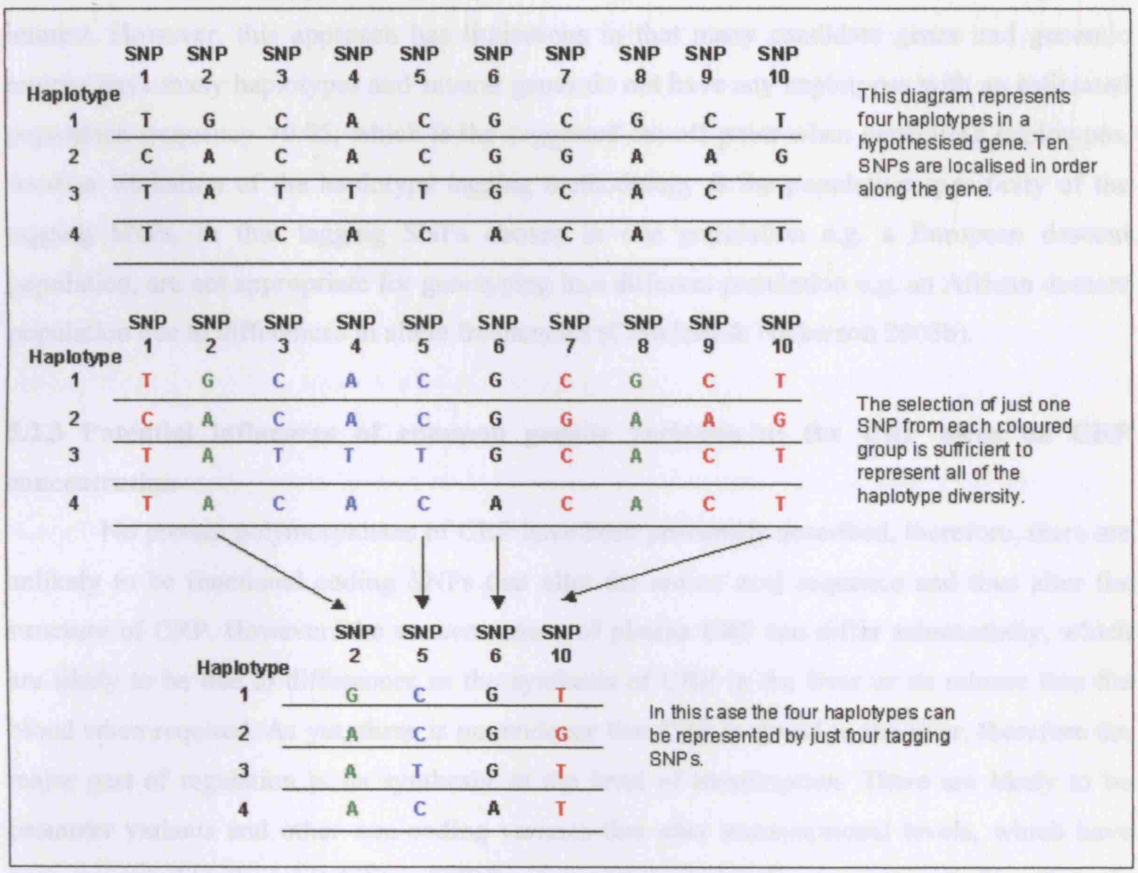
In most cases, the non-functional SNP is more likely to be typed than the functional SNP since there are more non-functional SNP present in the gene. Therefore, the SNP being typed may be a marker for the SNP that is not typed and may be functional. How well a typed

SNP captures a non-typed SNP depends on linkage disequilibrium (LD), on how tight the correlation is between the SNPs. LD can be measured using the D' statistic and the r^2 statistic. D' is obtained from D, the basic pairwise disequilibrium coefficient that corresponds to the difference between the probabilities of observing the alleles independently in the population. The r^2 statistic measures the correlation of alleles at two sites taking into account differences in allele frequency between two loci.

Generally, the LD is higher when SNPs are physically close to each other, because the chance of SNPs being separated at meiotic division is reduced. It was initially thought that the LD would fall away the further the distance was between two SNPs (Kruglyak 1999). However, it is now known that this is not the case, and the human genome is organised as blocks of high LD within which SNPs are often highly correlated with each other (Daly *et al.* 2001; Jeffreys *et al.* 2001; Gabriel *et al.* 2002).

This allows the potential for typing a subset of SNPs to capture the variation at any region as shown in Figure 5.2. If there was no LD between any pairs of SNPs in the hypothetical gene, in theory, there could be 2^{10} haplotypes. However, because groups of SNPs are in LD (as shown in the same colour), only four common haplotypes are observed and can be typed by any of the SNPs from each colour group. This essentially underlies the principle of haplotype tagging SNPs, and is useful for the development of a minimal set of SNPs that could be used for large-scale genotyping of similar populations.

Figure 5.2. Illustration of the use of LD knowledge between SNPs to generate tagging SNPs.



When information on SNPs is obtained from unrelated subjects, it is not possible to unambiguously assign a haplotype for an individual who is heterozygous at more than one locus because the linkage phase is unknown. Therefore, mathematical algorithms and programs have been developed that allow haplotypes to be inferred on the basis of probability of which is the likely haplotype using information from other individuals from the same data set who are homozygous for the SNPs being considered. These programs have also allowed the selection of tagging SNPs to adequately capture variation to be optimised.

Although family pedigrees and trio information eliminate some of the ambiguity of assigning alleles to chromosomes, this approach is very time-consuming and expensive. An alternative method is to use statistical inference to determine haplotypes from genotype data in large population-based studies of unrelated subjects. As the phase is not known in heterozygous subjects, it has to be inferred by parsimony, maximum likelihood or Bayesian methods (Crawford & Nickerson 2005a).

The use of haplotype tagging SNPs can greatly reduce the number of polymorphisms that would need to be typed to best represent the genetic variation of the candidate gene of interest. However, this approach has limitations in that many candidate genes and genomic regions have many haplotypes and several genes do not have any haplotypes with an estimated population frequency >0.05, which is the suggested cut-off point when generating haplotypes. Another limitation of the haplotype tagging methodology is the population specificity of the tagging SNPs, in that tagging SNPs chosen in one population e.g. a European descent population, are not appropriate for genotyping in a different population e.g. an African descent population due to differences in allele frequencies (Crawford & Nickerson 2005b).

5.2.3 Potential influences of common genetic variation at the CRP locus on CRP concentration

No protein polymorphisms of CRP have been previously described, therefore, there are unlikely to be functional coding SNPs that alter the amino acid sequence and thus alter the structure of CRP. However, the concentrations of plasma CRP can differ substantially, which are likely to be due to differences in the synthesis of CRP in the liver or its release into the blood when required. As yet, there is no evidence that CRP is stored in the liver; therefore the major part of regulation is its synthesis, at the level of transcription. There are likely to be promoter variants and other non-coding variants that alter transcriptional levels, which have been suggested by the heritability of CRP values and limited early association studies of single SNPs with CRP concentration. The aims of this chapter therefore, were to identify all reported SNPs in the CRP gene and generate a consensus map, and to examine reported pairwise LD between SNPs in order to select tagging SNPs, under the assumption that the tagging SNPs are transportable across populations of similar ancestry.

5.3 Methods

5.3.1 SNP databases

Several databases were searched to find information on the polymorphisms in the CRP gene and its flanking regions. The NCBI RefSeq collection contains non-redundant sequences including genomic DNA, gene transcripts (RNA), and protein products. The complete CRP gene sequence and its pseudogene sequence (accession number AF442818) are available on this site (http://www.ncbi.nlm.nih.gov/SNP/ snp_ref.cgi?locusId=1401), and gives details of confirmed and suspected SNPs within the genomic region of CRP.

The NHLBI Programs for Genomic Applications SeattleSNPS database (http://pga.gs.washington.edu/data/crp/) uses ~1Kb overlapping amplicons for dense re-sequencing and then utilises sequences from the NCBI RefSeq database to map SNPs onto the

gene. The HapMap Consortium database (http://www.hapmap.org/) is a culmination of information gathered from genotyping SNPs across the human genome. A total of four geographical populations have been genotyped in 270 individuals. These are 30 trios of Yoruba people from Ibadan, Nigeria (YRI), 45 unrelated Japanese people from Tokyo (JPT), 45 unrelated Chinese from Beijing (CHB), and 30 CEPH trios (both parents and one child) (CEU).

5.3.2 LD analysis

Using validated genotype data (http://pga.gs.washington.edu/gty_data/ crp/) for nine polymorphisms present in European individuals spanning CRP in 23 CEPH individuals, the LD architecture of CRP was analysed, utilising the software package TagIT v. 3.0 (http://www.genome.duke.edu/resources/computation/software). This program employs a linkage disequilibrium statistic known as the haplotype r^2 statistic that measures the correlation of alleles at two sites taking into account differences in allele frequencies between two loci and is more stringent than the D' measures of LD. D' is obtained from D, the basic pairwise disequilibrium coefficient that corresponds to the difference between the probabilities of observing the alleles independently in the population. Measures based on r^2 are advantageous as they have more reliable sample properties than D' for low allele frequencies, and it is less biased than D' when using small sample sizes. r^2 values greater than 0.8 were used to assess the LD structure.

The SeattleSNPs database also utilises a selection algorithm that is based on the LD statistic r^2 known as LDSelect (Carlson *et al.* 2004). This algorithm selects a subset of variants that efficiently describe all common patterns of variation in a gene, based on the minor allele frequency (MAF) of a SNP and the minimum level of association between genotyped and non-genotyped SNPs. LDSelect can therefore identify bins of SNPs whereby one tagSNP per bin can be genotyped that is representative of all the SNPs in that bin. Binning criteria for tagSNP selection are a MAF cutoff of 5% and an r^2 threshold of 0.65. The PHASE (v 2.0) haplotype inference program within the SeattleSNPs site allows all common haplotypes present in a particular population to be inferred, based on an r^2 greater than 0.64 between SNPs and a MAF greater than 5%.

The HapMap database has a program Tagger within their Haploview software package (http://www.hapmap.org/). This program allows LD and haplotype block analysis and haplotype population frequency estimation using D' and r^2 LD statistics. The minor allele frequency cut-off can be varied, as can the minimum values for D' and r^2. The LD between SNPs can be viewed as an LD plot, whereby blocks of LD are shown, and allow haplotypes to be inferred.

5.3.3 Promoter variation analysis

The MatInspector program (http://www.genomatix.de/products/MatInspector/index.html) was also used to examine if any polymorphisms in the promoter region were present in consensus sequences for transcription-factor binding, and whether base changes resulted in creation or loss of binding sites. MatInspector is a software tool that utilises a large library of matrix descriptions for transcription-factor binding sites to locate matches in DNA sequences. It assigns a quality rating to matches and therefore allows filtering and selection of matches (Quandt *et al.* 1995). The CRP promoter sequence taken from the SeattleSNPs database (~1050bp) was entered into the program in a FASTA format and a list of transcription-factor binding sites was generated. First the sites containing the major allele of polymorphisms were examined for creation or loss of binding sites, then the minor allele was examined in the same way.

5.4 Results

5.4.1 Polymorphisms in the CRP gene

SNPs reported in the NCBI SNP database were compared with those reported by the HapMap and SeattleSNPS databases to determine which polymorphisms were likely to be real, and which may have occurred through sequencing errors. A list of likely valid SNPs in the genomic and flanking regions of CRP was thus generated (see Table 5.2). No additional SNPs were identified at this locus from the HapMap database, which was published in the latter stages of this work.

Table 5.2. SNPs present in the CRP gene (taken from the NCBI SNP, SeattleSNPs, and HapMap databases), and their allele frequencies.

NCBI SNP	PGA SNP	Location	Allele 1	AD pop	ED pop	Allele 2	AD pop	ED pop	Allele 3	AD pop	ED pop
rs3093058	790	-936b	T	0.24	0.00	A	0.76	1.00			
rs3093059	969	-757b	C	0.35	0.07	T	0.65	0.93			
rs3093060	970	-756b	A	0.02	0.00	G	0.98	1.00			
rs2794521	1009	-717b	G	0.07	0.35	A	0.93	0.65			
rs3093061	1123	-603b	G	0.23	0.00	A	0.78	1.00			
rs3093062	1421	-305b	A	0.24	0.00	G	0.76	1.00			
rs3091244	1440	-286b	A	0.33	0.05	T	0.41	0.33	C	0.26	0.62
rs3093063	1632	-94b	T	0.02	0.00	C	0.98	1.00			
rs3093064	1724	-2b	T	0.04	0.00	C	0.96	1.00			
rs1417938	1919	+194b	T	0.17	0.26	A	0.83	0.74			
rs1800947	2667	+1059b	C	0.02	0.05	G	0.98	0.95			
rs3093065	2892	+1322b	T	0.02	0.00	C	0.98	1.00			
rs3093066	3006	+1436b	A	0.32	0.00	C	0.68	1.00			
rs1130864	3014	+1444b	T	0.14	0.30	C	0.86	0.70			
rs3093067	3577	+2007b	C	0.04	0.00	T	0.96	1.00			
rs1205	3872	+2302b	A	0.14	0.25	G	0.86	0.75			
rs2808631	4069	+2489b	G	0.11	0.00	A	0.89	1.00			
rs3093080	4362	+2792b	A	0.00	0.04	T	1.00	0.96			
rs3093068	4741	+3171b	G	0.31	0.07	C	0.69	0.93			
rs3093069	4974	+3404b	G	0.24	0.00	T	0.76	1.00			
rs2808630	5237	+3667b	G	0.06	0.36	A	0.94	0.64			
rs3093070	5288	+3718b	C	0.02	0.00	A	0.98	1.00			
rs3093071	5606	+4036b	A	0.02	0.02	C	0.98	0.98			
rs3093072	5676	+4106b	T	0.02	0.00	G	0.98	1.00			
rs3093073	5728	+4158b	C	0.02	0.00	T	0.98	1.00			
rs3093074	6150	+4580b	-	0.24	0.00	+	0.76	1.00			
rs3093075	6192	+4622b	A	0.34	0.07	C	0.66	0.93			
rs3093076	6333	+4763b	-	0.25	0.00	+	0.75	1.00			
rs3093077	6469	+4899b	G	0.39	0.07	T	0.61	0.93			
rs3093078	6556	+4986b	T	0.05	0.00	C	0.95	1.00			
rs3093079	6677	+5107b	T	0.02	0.00	C	0.98	1.00			

AD: African Descent; ED: European Descent; pop: population.

Diallelic insertion/deletion polymorphisms are represented as:

+ : insertion allele; - : deletion allele

The + (inserted) allele for each site is listed below:

rs3093074/6150: T

rs3093076/6333: GAT

SNPs: = flanking, = intron, = synom, = 3'UTR;

The NCBI SNP (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=1401) database had a total number of 27 SNPs. Out of these SNPs, one was a non-coding variant in exon 2 (see Table 5.3). Ten SNPs were present in the 5' upstream promoter region, two in the intron, and the rest were in the 3' UTR and further downstream. One of the intronic SNPs was a biallelic GT repeat, which can be repeated between 15 and 24 times. The PGA SeattleSNPs database (http://pga.gs.washington. edu/data/crp/) contained a total of 31 variants, including some that were exclusive to subjects of African descent (discussed in more detail in Chapter 9). These included nine promoter and 5' upstream variant, one intronic variant, a synonymous variant in exon 2, and twenty SNPs in the 3'UTR and downstream region. Out of the SNPs in the 3' downstream region, there were two insertion/deletion polymorphisms (see Table 5.3). In the HapMap database, there were 25 SNPs that were present in either the Nigerian, Japanese, Chinese or European subjects. These included six promoter variants, one intronic SNP, a non-coding SNP in exon 2 and seventeen SNPs in the 3'UTR and downstream region (see Table 5.3).

Table 5.3. Polymorphisms in the genomic and flanking regions of the CRP gene, taken from the NCBI SNP, SeattleSNPs, and HapMap databases. (Polymorphisms include single base changes with a frequency ≥ 0.01).

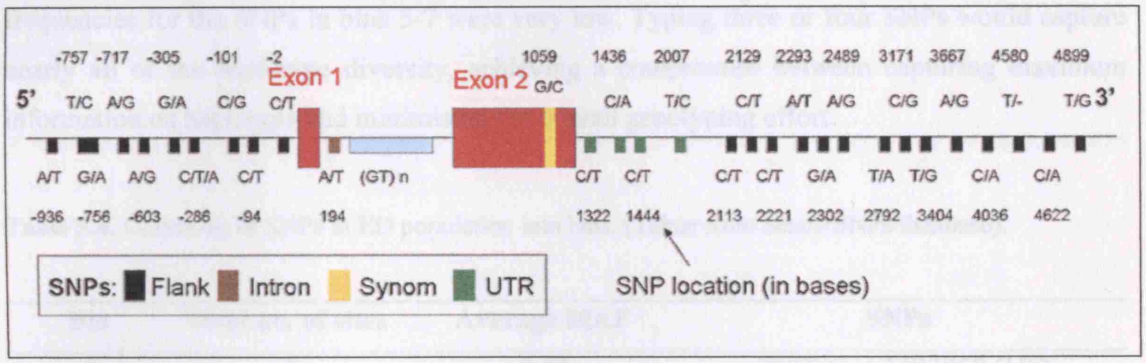
SNP ID (rs number)	SNP	NCBI SNP Location	SeattleSNPs Location	HapMap Location
3093058	A/T	5' Flanking	5' Flanking	5' Flanking
3093059	T/C	5' Flanking	5' Flanking	5' Flanking
3093060	G/A	5' Flanking	5' Flanking	5' Flanking
2794521	C/T	5' Flanking	5' Flanking	/
3093061	A/G	5' Flanking	5' Flanking	/
3093062	A/G	5' Flanking	5' Flanking	/
3091244	C/A/T	5' Flanking	5' Flanking	/
3122011	C/G	5' Flanking	/	5' Flanking
3093063	C/T	5' Flanking	5' Flanking	5' Flanking
3093064	C/T	5' Flanking	5' Flanking	5' Flanking
1417938	A/T	Intron	Intron	Intron
3138528	$(GT)_R$	Intron	/	/
1800947	G/C	Exon 2	Exon 2	Exon 2
3093065	C/T	3' UTR	3' UTR	3' UTR
3093066	C/A	3' UTR	3' UTR	3' UTR
1130864	C/T	3' UTR	3' UTR	3' UTR
1050031	C/A	3' UTR	/	/
3093067	T/C	3' UTR	3' UTR	/
3116637	C/T	3' UTR	/	3' UTR
6413465	C/T	3' Flanking	/	/
6413466	C/T	3' Flanking	/	3' Flanking
3116638	C/T	3' Flanking	/	3' Flanking
3116639	C/T	3' Flanking	/	/
3116640	C/T	3' Flanking	/	/
6413467	A/T	3' Flanking	/	/
1205	G/A	3' Flanking	3' Flanking	3' Flanking
2808631	A/G	3' Flanking	3' Flanking	/
3093080	T/A	/	3' Flanking	3' Flanking
3093068	C/G	/	3' Flanking	/
3093069	T/G	/	3' Flanking	3' Flanking
2808630	A/G	/	3' Flanking	3' Flanking
3093070	A/C	/	3' Flanking	3' Flanking
3093071	C/A	/	3' Flanking	3' Flanking
3093072	G/T	/	3' Flanking	3' Flanking
3093073	T/C	/	3' Flanking	3' Flanking
3093074	T/-	/	3' Flanking	/
3093075	C/A	/	3' Flanking	3' Flanking
3093076	GAT/-	/	3' Flanking	/
3093077	T/G	/	3' Flanking	3' Flanking
3093078	C/T	/	3' Flanking	/
3093079	C/T	/	3' Flanking	3' Flanking

/ = SNP not reported in this database.

By compiling a list of the SNPs and their location in the CRP gene, it was possible to construct a consensus SNP map of the gene, which included the polymorphisms in the flanking

5' and 3' ends, as well as in the genomic region. Polymorphisms that were detailed in at least two databases were included in the genetic map (see Figure 5.3). A total of 31 gene variants were identified, including 10 5' flanking SNPs, 2 intronic variants, 1 synonymous coding polymorphism (+1059), 4 3'UTR SNPs, and 14 3' flanking variants. Out of these variants, 29 were reported in African populations and 12 were reported in European populations (see Chapter 9 for more details).

Figure 5.3. Genetic map of CRP showing locations of polymorphisms within the genomic region and flanking regions. Location range is from –936bases to +4899bases.



The total length of the CRP gene and flanking regions included in the genetic map was around 6050bp. The majority of the polymorphisms were found in the 3' UTR and flanking region. Only one SNP was identified in the coding region of exon 2, which was a synonymous base change from G to C. As the gene contained no common non-synonymous variants identified so far, it implied that CRP is a homogenous protein so that differences, if any, in its action should be governed solely by its level. However, very rare variants may exist in some populations that affect the amino acid sequence, and thus may alter its properties and its action. The (GT) repeat in the intron (rs3138528) was not present in the SeattleSNPs and HapMap databases, even though it was one of the first polymorphisms to be identified in the CRP gene (Goldman *et al.* 1987), and has since been confirmed in other studies (Weber *et al.* 1990; Szalai *et al.* 2002).

5.4.2 Linkage disequilibrium in CRP

Analysis of LD between polymorphic sites in CRP was performed using data from individuals in the SeattleSNPs database to identify clusters of highly correlated sites based on the r^2 LD statistic. SNP information was used from this database since it was the only one to type SNPs in several unrelated subjects. SNPs were separated into sets of tightly correlated markers, referred to as bins. In the African descent (AD) population, there were 18 bins, and in

the European descent (ED) population there were 7 bins. This implied that a smaller set of SNPs was required to obtain information on all the SNPs in an ED population, compared to an AD population.

Linkage disequilibrium and tagging SNPs in European populations

The largest bin from the ED population (Bin 1, see Table 5.4) contained four SNPs. Typing any one of these SNPs provided near unequivocal information about the other three SNPs in the same bin. Similarly, the two next largest bins each contained two SNPs so typing one SNP from each of these bins would provide information about the other SNP in that bin. Although there were seven bins for the ED population, and therefore eight haplotypes, the allele frequencies for the SNPs in bins 5-7 were very low. Typing three or four SNPs would capture nearly all of the haplotype diversity, achieving a compromise between capturing maximum information on haplotype and minimising the overall genotyping effort.

Table 5.4. Clustering of SNPs in ED population into bins. (Taken from SeattleSNPs database).

Bin	Total no. of sites	Average MAF	SNPs
1	4	0.07	-757T/C; rs3093059 (UF)
			+3171C/G; rs3093068 (RF)
			+4622C/A; rs3093075 (UF)
			+4899T/G; rs3093077 (UF)
2	3	0.32	-286C/T/A; rs3091244 (UF)
			+194A/T; rs1417938 (UI)
			+1444C/T; rs1130864 (UU)
3	2	0.36	-717A/G; rs2794521 (RF)
			+3667A/G; rs2808630 (UF)
4	1	0.25	+2302G/A; rs1205 (RF)
5	1	0.05	+1059G/C; rs1800947 (US)
6	1	0.02	+4036C/A; rs3093071 (UF)
7	1	0.04	+2792T/A; rs3093080 (UF)

MAF: minor allele frequency.

The first position of the code indicates whether the sequence context is: (U)nique sequence or (R)epeat containing sequence.

The second position in the code provides information on the genomic context: (F)lanking region, 5' or 3' (U)TR, (I)ntron, (S)ynonymous cSNP, or (N)onsynonymous cSNP.

The CRP genotype data for 10 typed SNPs with a minor allele frequency greater than 5% among the 23 European CEPH individuals from the SeattleSNPs database were entered into the TagIT program and used to generate a SNPhap output, which gives the probabilities for each haplotype occurring. In the absence of LD, 10 SNPs would produce 2^{10} haplotypes, but because of the extensive LD at this locus (see Table 5.5), only four common haplotypes (frequency >0.05) were inferred with estimated frequencies 0.065 to 0.348 (see Table 5.6). Based on the r^2 statistic, 8 trios of potential tagging SNPs captured perfectly the genetic diversity for the inputted data (see Table 5.7). One trio, comprising SNPs +1444C/T (rs1130864), +2302G/A (rs1205) and +4899T/G (rs3093077) was selected to capture the variation in the CRP gene as shown in Figure 5.4, based on proximity to other polymorphisms and feasibility of genotyping using the TaqMan methodology (see Chapter 4).

Table 5.5. r^2 statistic showing LD between SNPs in the CRP gene in a European population generated from the TagIT program using 23 European descent samples.

SNP	-757	-717	+194	+1444	+2302	+3171	+3667	+4622	+4899
-757	1	0.0372	0.0305	0.0305	0.0275	1	0.0372	1	1
-717	0.0372	1	0.2333	0.2333	0.2101	0.0372	1	0.0372	0.0372
+194	0.0305	0.2333	1	1	0.1723	0.0305	0.2333	0.0305	0.0305
+1444	0.0305	0.2333	1	1	0.1723	0.0305	0.2333	0.0305	0.0305
+2302	0.0275	0.2101	0.1723	0.1723	1	0.0275	0.2101	0.0275	0.0275
+3171	1	0.0372	0.0305	0.0305	0.0275	1	0.0372	1	1
+3667	0.0372	1	0.2333	0.2333	0.2101	0.0372	1	0.0372	0.0372
+4622	1	0.0372	0.0305	0.0305	0.0275	1	0.0372	1	1
+4899	1	0.0372	0.0305	0.0305	0.0275	1	0.0372	1	1

The –286C/T/A polymorphism was not included in this analysis as it was triallelic.

Table 5.6. SNPhap output showing probabilities of haplotypes from the TagIT program based on 10 SNPs in 23 European descent samples.

	-757	-717	+194	+1059	+1444	+2302	+3171	+3667	+4622	+4899	Frequency
Haplotype 1	2	1	1	1	1	1	2	1	2	2	0.0652
Haplotype 2	1	2	1	1	1	1	1	2	1	1	0.3478
Haplotype 3	1	1	2	1	2	1	1	1	1	1	0.3044
Haplotype 4	1	1	1	2	1	2	1	1	1	1	0.2432
Haplotype 5	1	1	1	2	1	1	1	1	1	1	0.0394

Table 5.7. Trios of potential tagging SNPs that all capture most of the variation in the CRP gene in a European population.

Trio	SNP 1	SNP 2	SNP 3
Trio 1	+1444	+2302	+4622
Trio 2	+1444	+2302	+4899
Trio 3	+1444	+3171	+3667
Trio 4	+1444	+3667	+4622
Trio 5	+1444	+3667	+4899
Trio 6	+2302	+3171	+3667
Trio 7	+2302	+3667	+4622
Trio 8	+2302	+3667	+4899

Figure 5.4. Genetic map of CRP showing locations of the haplotype tagging SNPs within the CRP gene in European populations.

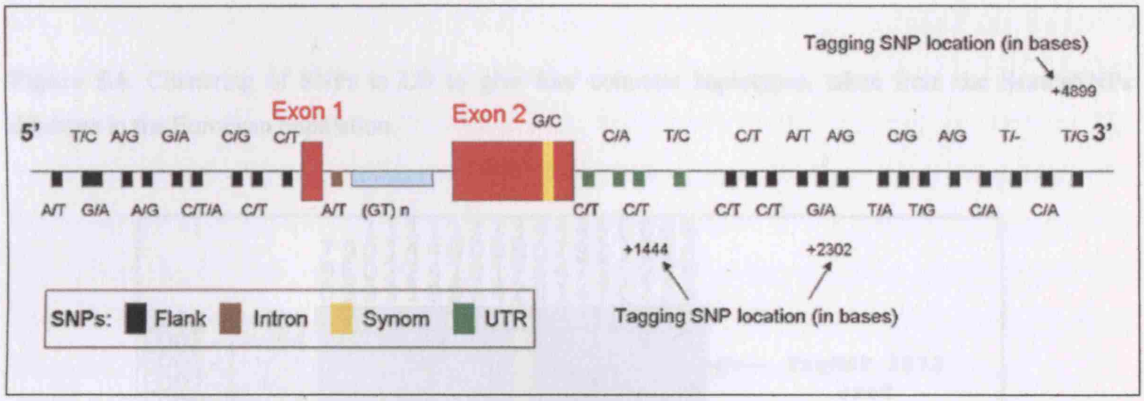
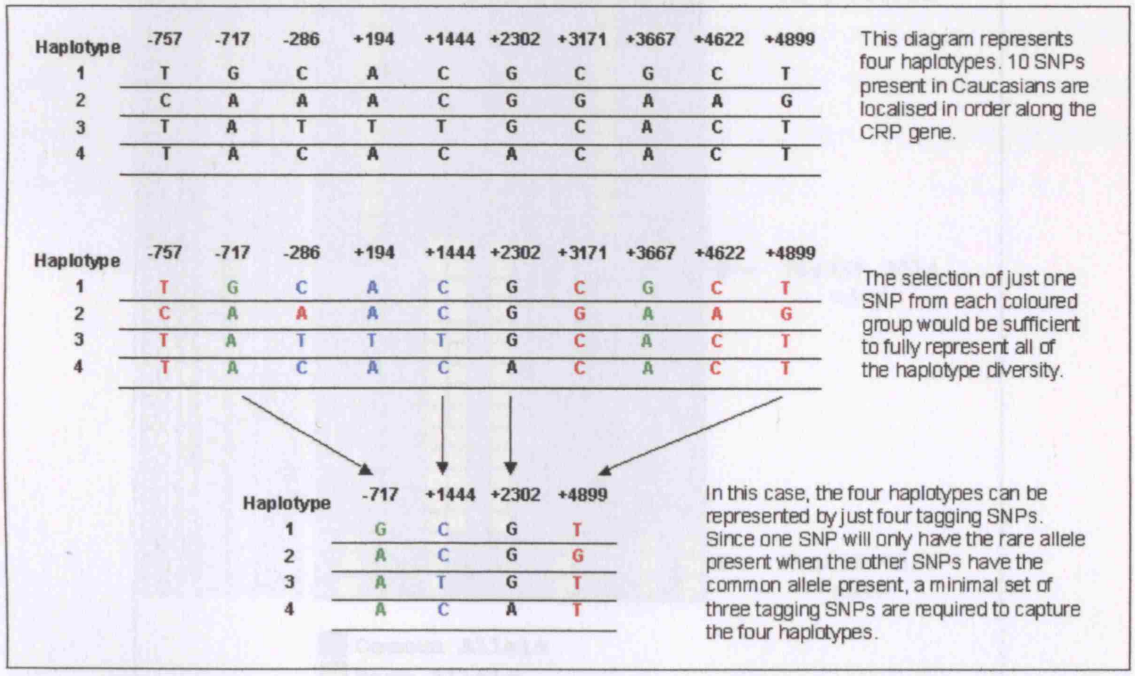


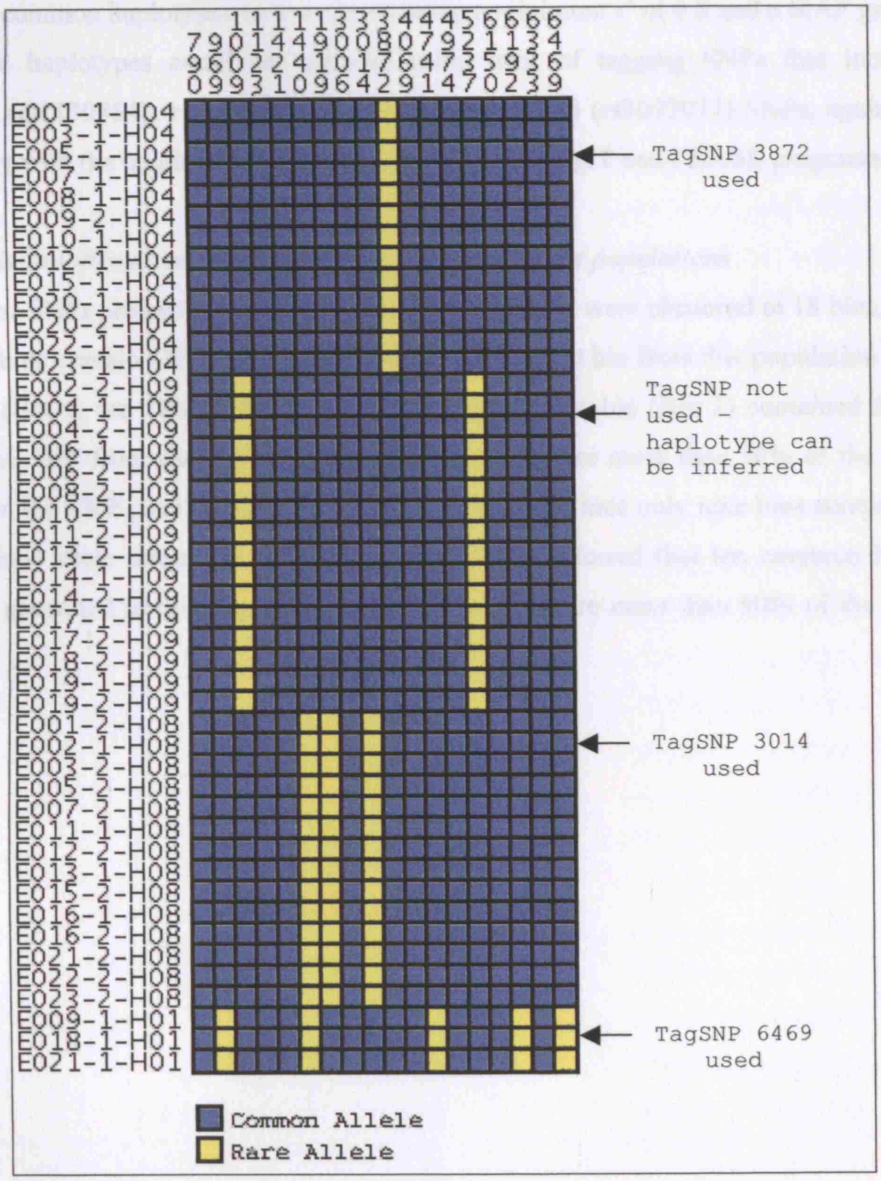
Figure 5.5. Overview of the haplotype tagging approach to select tagging SNPs in the CRP gene in European populations.



The SNP data from the 23 individuals in the SeattleSNPs database was then inputted into PHASE (v 2.0) haplotype inference program (used on the SeattleSNPs site) to determine the haplotype data for each individual (Stephens *et al.* 2001b). This was done to check whether the selection of tagging SNPs from the TagIT program was adequate to capture most of the variation in the CRP gene. From the PHASE, it was determined that by genotyping just three

SNPs, rs1130864 (PGA3014, +1444C/T), rs1205 (PGA3872, +2302G/A) and rs3093077 (PGA6469, +4899T/G) in a Caucasian population to generate 4 common haplotypes as shown in Figure 5.6, the majority of the information of the variants in the gene would be obtained.

Figure 5.6. Clustering of SNPs in LD to give four common haplotypes, taken from the SeattleSNPs database in the European population.



The rows represent different individuals of European descent; the columns represent SNPs in the gene from 5' to 3'. Arrows identify which SNPs were subsequently chosen as tagging SNPs.

The PHASE program showed that the 969 (rs3093059), 4741 (rs3093068), 6192 (rs3093075) and 6469 (rs3093077) SNPs were in LD, the 1440 (rs3091244), 1919 (rs1417938) and 3014 (rs1130864) SNPs were in LD, and the 1009 (rs2794521) and 5237 (rs2808630) SNPs were in LD. This was consistent with the data generated from the TagIT program and therefore confirmed the choice of tagging SNPS selected from the TagIT program. Indeed, when the SNP data from the CEPH European population from the HapMap database were assessed for LD structure and generation of haplotypes in the program Tagger (http://www.hapmap.org/), the same four common haplotypes were inferred using a minimum r^2 of 0.8 and a MAF greater than 5%. These haplotypes could be captured using trios of tagging SNPs that included the +1444C/T (rs1130864), +2302G/A (rs1205) and +4899T/G (rs3093077) SNPs, again showing consistency with the tagging SNP set generated from the TagIT and PHASE programs.

Linkage disequilibrium and tagging SNPs in African descent populations

The SNPs present in the African descent population were clustered in 18 bins, however, only nine bins contain SNPS with a MAF >0.05. The largest bin from this population contained six SNPs (Bin 1, see Table 5.8). Similarly, the next largest bin (Bin 2) contained five SNPs. Typing one SNP from each of these two bins could capture more than 50% of the haplotype diversity in the CRP gene in African descent populations. Since only nine bins contained SNPs with a minor allele frequency greater than 5%, it was inferred that ten common haplotypes would be generated in this population, which would capture more than 90% of the haplotype diversity.

Table 5.8. Clustering of SNPs in AD population into bins. (Taken from SeattleSNPs database).

Bin	Total no. of sites	Average MAF	SNPs
1	6	0.31	-757T/C; rs3093059 (UF)
			+1436C/A; rs3093066 (UU)
			+3404T/G; rs3093069 (UF)
			+4580T/-; rs3093074 (UF)
			+4622C/A; rs3093075 (UF)
			+4899T/G; rs3093077 (UF)
2	5	0.35	-936A/T; rs3093058 (UF)
			-603A/G; rs3093061 (RF)
			-305G/A; rs3093062 (UF)
			+3171C/G; rs3093068 (RF)
			+4763GAT/-; rs3093076 (RF)
3	2	0.06	-717A/G; rs2794521 (RF)
			+3667A/G; rs2808630 (UF)
4	2	0.02	-756G/A; rs3093060 (UF)
			+4106G/T; rs3093072 (UF)
5	2	0.04	-2C/T; rs3093064 (UF)
			+2007T/C; rs3093067 (UU)
6	1	0.05	+4986C/T; rs3093078 (UF)
7	1	0.14	+2302G/A; rs1205 (RF)
8	1	0.14	+1444C/T; rs1130864 (UU)
9	1	0.17	+194A/T; rs1417938 (UI)
10	1	0.02	+1059G/C; rs1800947 (US)
11	1	0.11	+2489A/G; rs2808631 (UF)
12	1	0.59	-286C/T/A; rs3091244 (UF)
13	1	0.02	-94C/T; rs3093063 (UF)
14	1	0.02	+1322C/T; rs3093065 (UU)
15	1	0.31	+2792T/A; rs3093080 (UF)
16	1	0.02	+3718A/C; rs3093070 (UF)
17	1	0.02	+4036C/A; rs3093071 (UF)
18	1	0.02	+4158T/C; rs3093073 (UF)

MAF: minor allele frequency.

The first position of the code indicates whether the sequence context is: (U)nique sequence or (R)epeat containing sequence.

The second position in the code provides information on the genomic context: (F)lanking region, 5' or 3' (U)TR, (I)ntron, (S)ynonymous cSNP, or (N)onsynonymous cSNP.

The LDSelect program was then used to assess the pairwise LD between SNPs in the African descent population using data on this population from the SeattleSNPs database. A total of 16 SNPs with a MAF >0.05 were evaluated using the r^2 statistic with a minimum value of 0.64. In the absence of LD, 16 SNPs would produce 2^{16} haplotypes, but because of the extensive LD at this locus, far fewer haplotypes were seen (see Table 5.9).

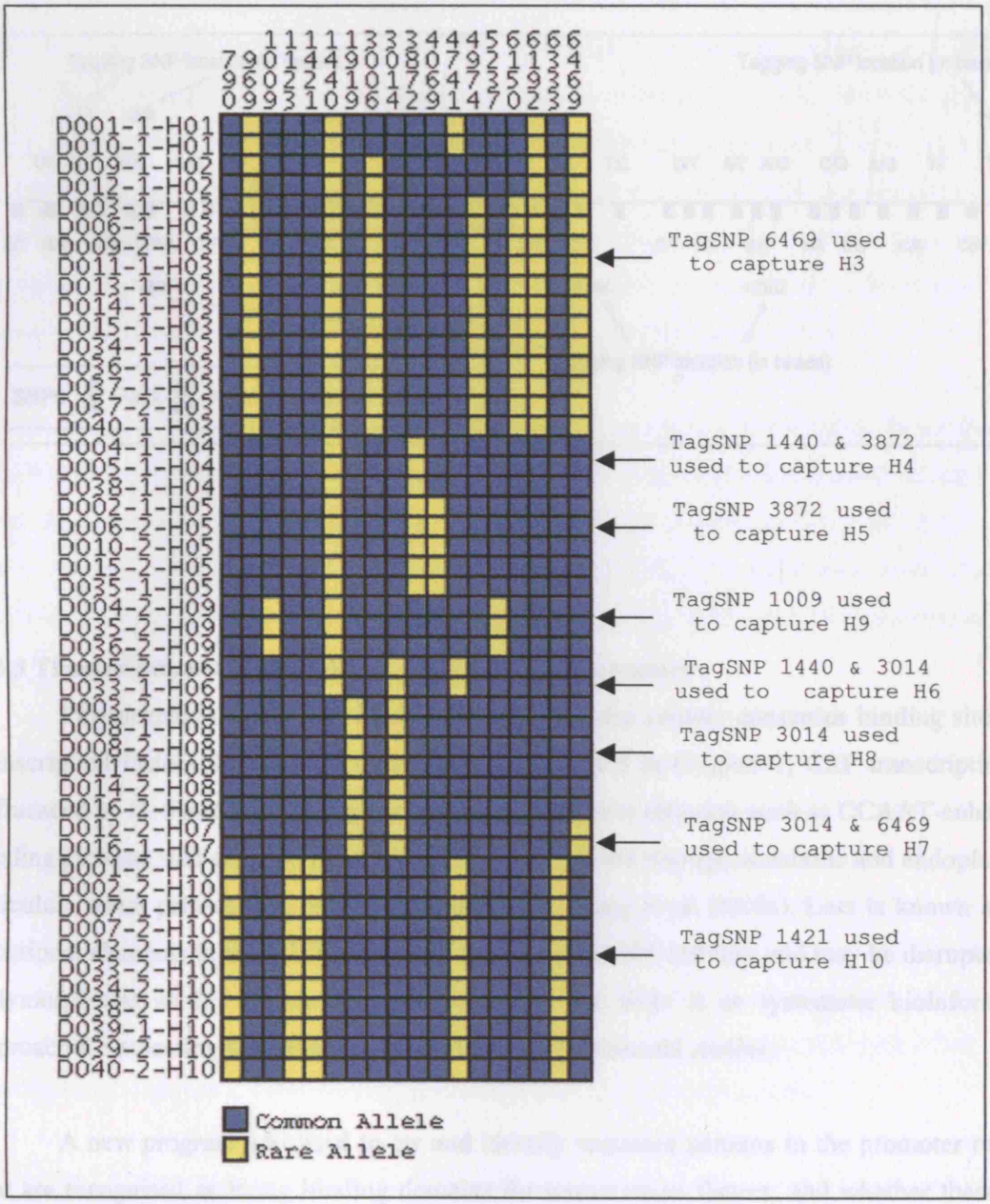
The PHASE (v 2.0) haplotype inference program from the SeattleSNPS database was also used to generate tagging SNPS in an African descent population. From the SeattleSNPS database, ten haplotypes were inferred, as shown in Figure 5.7. The PHASE program determined that typing a minimum of six SNPs, rs2794521 (PGA1009, -717A/G), rs3093062 (PGA1421, -305G/A), rs3091244 (PGA1440, -286C/T/A), rs1130864 (PGA3014, +1444C/T), rs1205 (PGA3872, +2302G/A) and rs3093077 (PGA6469, +4899T/G) in an African population would capture the majority of the haplotype diversity in the CRP gene (see Figure 5.8).

The Tagger program from the HapMap site was also used to examine tagging SNPs. Data on SNPs in the Nigerian African population from the HapMap database were entered into the program using a MAF>0.05 and $r^2 \geq 0.8$. A total of 15 SNPs have been genotyped in this population and have a minor allele frequency greater than 5%. The –286C/T/A polymorphism has not been typed and could therefore not contribute to the tagging SNP set. Using the Tagger program, two SNPs were selected as tagging SNPS in addition to the ones selected by PHASE (rs3093066 and rs3093069), that also generated ten haplotypes. These results are consistent with the tagging SNPs and haplotypes generated from the programs on the SeattleSNPs site.

Table 5.9. r^2 statistic showing LD between 16 SNPs in the CRP gene in an African descent population generated from the LDSelect program using 24 African descent samples.

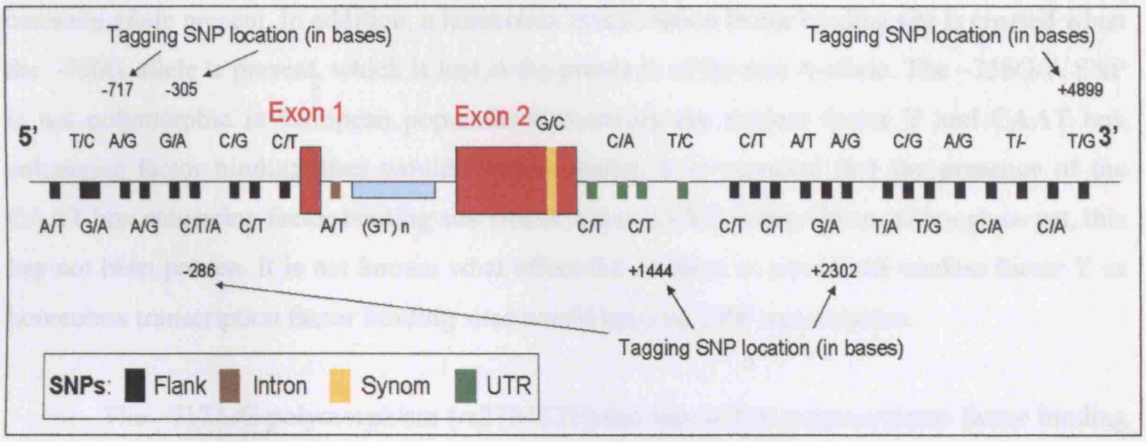
SNP	-936	-757	-717	-603	-305	-286	+194	+1436	+1444	+2007	+2302	+2489	+3171	+3404	+3667	+4899
-936	1	0.17	0.02	1	1	0.41	0.05	0.12	0.03	0.02	0.05	0.04	0.70	0.10	0.02	0.14
-757	0.17	1	0.04	0.17	0.17	0.43	0.09	0.82	0.10	0.02	0.10	0.07	0.14	0.64	0.04	0.85
-717	0.02	0.04	1	0.02	0.02	0.06	0.01	0.03	0.01	0.00	0.14	0.01	0.03	0.02	1	0.03
-603	1	0.17	0.02	1	1	0.48	0.04	0.14	0.05	0.01	0.04	0.03	0.60	0.10	0.02	0.15
-305	1	0.17	0.02	1	1	0.34	0.07	0.14	0.05	0.01	0.04	0.03	0.72	0.10	0.02	0.15
-286	0.41	0.43	0.06	0.48	0.34	1	0.22	0.37	0.03	0.05	0.12	0.09	0.10	0.30	0.06	0.40
+194	0.05	0.09	0.01	0.04	0.07	0.22	1	0.00	0.63	0.22	0.03	0.02	0.01	0.07	0.01	0.03
+1436	0.12	0.82	0.03	0.14	0.14	0.37	0.00	1	0.00	0.02	0.09	0.07	0.20	0.79	0.03	0.80
+1444	0.03	0.10	0.01	0.05	0.05	0.03	0.63	0.00	1	0.15	0.03	0.03	0.00	0.02	0.01	0.03
+2007	0.02	0.02	0.00	0.01	0.01	0.05	0.22	0.02	0.15	1	0.01	0.01	0.02	0.01	0.00	0.02
+2302	0.05	0.10	0.14	0.04	0.04	0.12	0.03	0.09	0.03	0.01	1	0.63	0.07	0.06	0.15	0.09
+2489	0.04	0.07	0.01	0.03	0.03	0.09	0.02	0.07	0.03	0.01	0.63	1	0.05	0.03	0.01	0.07
+3171	0.70	0.14	0.03	0.60	0.72	0.10	0.01	0.20	0.00	0.02	0.07	0.05	1	0.14	0.03	0.14
+3404	0.10	0.64	0.02	0.10	0.10	0.30	0.07	0.79	0.02	0.01	0.06	0.03	0.14	1	0.02	0.67
+3667	0.02	0.04	1	0.02	0.02	0.06	0.01	0.03	0.01	0.00	0.15	0.01	0.03	0.02	1	0.03
+4899	0.14	0.85	0.03	0.15	0.15	0.40	0.03	0.80	0.03	0.02	0.09	0.07	0.14	0.67	0.03	1

Figure 5.7. Clustering of SNPs in LD to give eight common haplotypes, taken from the SeattleSNPs database in the African descent population.



The rows represent different individuals of African descent; the columns represent SNPs in the gene from 5' to 3'. Arrows identify which SNPs were subsequently chosen as tagging SNPs. Numbers with the prefix 'H' represent the haplotype number.

Figure 5.8. Genetic map of CRP showing locations of the haplotype tagging SNPs within the CRP gene in African descent populations.



5.4.3 Transcription-factor binding sites in the CRP promoter

Polymorphic variation in the 5' UTR of genes may contain consensus binding sites for transcription factors that modify expression. As outlined in Chapter 1, CRP transcription is influenced by IL-6 and IL-1 through acute phase responsive elements such as CCAAT-enhancer binding proteins, and also potentially by USF1 and CREBH through metabolic and endoplasmic reticulum stress pathways (Corre & Galibert 2005; Zhang *et al.* 2006a). Less is known about functional elements in the 3'UTR that are involved in mRNA stability and may be disrupted by polymorphisms in this region. This is because as yet, there is no systematic bioinformatic approach for assessing this, apart from conducting experimental studies.

A new program was used to try and identify sequence patterns in the promoter region that are recognised as being binding domains for transcription factors, and whether these are created or disrupted by polymorphisms in the CRP gene. The MatInspector program was executed using a 1050bp region of the promoter sequence, entered as a FASTA sequence. This region contains nine polymorphisms, the –936A/T (rs3093058), -757T/C (rs3093059), -756G/A (rs3093060), -717A/G (rs2794521), -603A/G (rs3093061), -305G/A (rs3093062), -286C/T/A (rs3091244), -94C/T (rs3093063) and –2C/T (rs3093064). Out of these SNPs, six polymorphisms were found to lie within transcription-factor binding sites, as shown in Figure 5.9.

The -757T/C (rs3093059) and -756G/A (rs3093060) polymorphisms have several transcription-factor binding sites located here, some of which are overlapping and require both

rare alleles, due to their proximity. When the rare alleles of both SNPs are present (-757C and –756A), a nuclear factor Y trancription factor binding site and CAAT box enhancing factor binding site are created. These binding sites are lost when one or both of these SNPs have the common allele present. In addition, a homeobox transcription factor binding site is created when the –756G-allele is present, which is lost in the presence of the rare A-allele. The –756G/A SNP is not polymorphic in European populations, therefore the nuclear factor Y and CAAT box enhancing factor binding sites would not be created. It is expected that the presence of the CAAT box enhancing factor binding site would increase CRP transcription, although as yet, this has not been proven. It is not known what effect the creation or loss of the nuclear factor Y or homeobox transcription factor binding sites would have on CRP transcription.

The –717A/G polymorphism (rs2794521) also lies within a transcription factor binding site. When the major A-allele is present, a heat shock factor binding site is created, which is lost when the rare G-allele is present. When the minor G-allele is present, an E-box N-Myc binding site is created, which is not seen in the presence of the A-allele. It is not known what effect the creation or loss of these transcription factor binding sites would have on CRP transcription.

The –305G/A polymorphism (rs3093062) is not present in European populations, where the G-allele would be continuously present. Since the presence of the G-allele creates E-box upstream stimulating factor (USF) and N-Myc binding sites, these binding sites would be continuously present in European populations. In other populations, where this site is polymorphic, such as African descent populations, when the rare A-allele is present, an acute myelogenous leukaemia factor binding site is created, and the USF and N-Myc binding sites are lost. It is expected that the presence of E-box transcription factor binding sites result in an increase in CRP transcription. However, it is not known what effect the presence of an acute myelogenous leukaemia factor binding site would have on CRP transcription.

The –286C/T/A (rs3091244) triallelic polymorphism also lies within trancription factor binding sites. When the T-allele is present, USF trancription factor and forkhead related transcription factor (FREAC) binding sites are created. These binding sites are lost in the presence of the C- or A-alleles. It is expected that the presence of E-box transcription factor binding sites result in an increase in CRP transcription. However, it is not known what effect the presence of a forkhead related transcription factor binding site would have on CRP transcription.

A recent functional study has been conducted by Szalai et al., investigating the effects of these two promoter polymorphisms on CRP concentrations. Both the –286T and –305G sites were found to encode E-boxes that were functional from electrophoretic mobility shift assay

(EMSA) studies. Healthy individuals that carried both the E-box binding sites (-286T and -305G) had the highest baseline CRP concentrations whereas individuals carrying the C- or A-allele at the –286 and the A-allele at the –305 sites had the lowest CRP concentrations (Szalai *et al.* 2005).

The –94C/T polymorphism (rs3093063) is not present in European populations, where the C-allele would be continuously present. Since the presence of the C-allele creates E-box upstream stimulating factor (USF) and MyoD binding sites, these binding sites would be continuously present in European populations. In other populations, where this site is polymorphic, such as African descent populations, when the rare A-allele is present, the USF and MyoD binding sites are lost. It is expected that the presence of E-box transcription factor binding sites result in an increase in CRP transcription. However, it is not known what effect the presence of a MyoD trancription factor binding site would have on CRP transcription.

Figure 5.9. Region of the CRP promoter analysed for alterations in transcription-factor binding sites due to polymorphisms. SNPs are highlighted in red and transcription-factor binding sites are highlighted according to whether a site is created or lost in the presence of the minor allele.

```
-975  GTGTGGAGGG ATTACTTGAA TCTTGTGAAT AGAGGAAAGA/t GTAGAATCAG (-936)
-925  ATTATCCTGA CTCCTGCCTG AAGCTTTACA TATTCAGAGA AAAATGTTGG
-875  AAGAAACTTT GATATAATGC TATGTCTGTG ATCAGGCACA CATTTTACTG
-825  GACTTTTACT GTCAGGGCCG TCATTTAGTG CCAAGATGTC TAGAGAGTTC
-775  TTAATAAGTG TACTCAATT/cG/a GCTGAGAAAA TGTGTCCATG CAAAAAACCA
           ECAT/NFY; PCAT/CAAT (minor allele of -757 & -756 SNPs)
                      HOXF/HOX1-3 (-756 SNP only)
-725  AACACCGCA/gT GTTCTCACTC ATAGATGGGA ATTGAACAAT GAGAATACTT
     HEAT/HSF1; EBOX/NMYC (-717 SNP)
-675  GGACACAGGA AGGGGAACAT CACACTCTGG GGACTGTTGT GGGGTGGGGG
-625  GAGGGGGGAG GGATAGCATT AGA/gAGATATA CCTAATGCTA AATGATGAGT (-603)
-575  TAATGGGTGC AGCACACCAG CATGGCACAT GTATACATAT GTAACTAACC
-525  TGCACATTGT GCACATGTAC CCTAAAACTT AAAGTATAAT AATAATAAAA
-475  AAATGTGTCC ATGGCTCTGG GAGGAGCATG TTTGTTTTCC TCATTTCCCA
-425  GTCTGTAAAT AAGCAAATTG AAAGGGGTTA GTGATAATGT CCATCTCCAG
-375  AAGCTGTCAG ATTTCCTTTG TCAAACTCTA TGATTTGGGC TGAAGTAGGT
-325  GTTGGAGAGG CAGCTACCAC G/aTGCACCCAG ATGGCCACTC/t/a GTTTAATATG
   EBOX/USF; EBOX/NMYC; HAML/AML (-305)  EBOX/USF; FKHD/FREAC2 (-286)
-275  TTACCATTTC CCATTATTTT CGCAGGATAG ATAGCCAAAG TGGAGCCCTG
-225  AGAGATTTCT TCATTTTTCC TGTCATAAAG AATTGGTAAT TCAGTAGTCA
-175  TAGGAGTTTG TAATAAATAA CTCACATTGA TTTCTCTGTT CTGAAATAAT
-125  TTTGCTTCCC CTCTTCCCGA AGCTCTGACA CC/tTGCCCCAA CAAGCAATGT
                                  MYOD; EBOX/USF (-94 SNP)
-75   TGGAAAATTA TTTACATAGT GGCGCAAACT CCCTTACTGC TTTGGATATA
-25   AATCCAGGCA GGAGGAGGTA GCTC/tTAAGGC AAGAGATCTA GGACTTCTAG (-2)
```

Key:

Binding sites lost due to presence of minor allele

Binding sites created due to presence of minor allele

ECAT/NFY	Nuclear factor Y
PCAT/CAAT	CAAT box enhancing factor
HOXF/HOX1-3	Homeobox transcription factors 1-3
EBOX/NMYC	N-Myc transcription factor
HEAT/HSF1	Heat shock factor 1
EBOX/USF	Upstream stimulating factor
HAML/AML	Human acute myelogenous leukaemia factor
FHKD/FREAC2	Forkhead related activator-2
MYOD	MyoD transcription factor

5.5 Discussion

Recent advances in sequencing projects such as the Human Genome project have shown that there are frequent common variants throughout the genome. Some of these are considered to underlie disease susceptibility or lead to variation in traits. Most polymorphisms throughout the genome are not functional, but they are useful to help map genetic determinants of disease susceptibility or intermediate traits, based on their ability to capture non-typed functional variants. How well a typed SNP captures a non-typed SNP depends on linkage disequilibrium (LD), on how tight the correlation is between the SNPs.

Earlier work in CRP-CHD association studies has been carried out utilising single polymorphisms, which may not capture the full range of variation at a locus. The development of SNP maps across the genome and dense re-sequencing efforts related to candidate genes, offers the possibility of a haplotype-based approach. At any given locus, the genotyping effort required in association studies may be decreased by identifying polymorphisms that are in LD with one another to generate a minimum set of SNPs (haplotype tagging SNPs) that capture haplotype diversity. Knowledge of generic variations in the CRP gene could facilitate Mendelian randomisation analysis of CRP and disease outcome, if SNPs can be identified that are associated with CRP concentration.

The aim of the work in this chapter was to identify all the SNPs in the CRP gene utilising public domain databases, examine the linkage disequilibrium (LD) between the identified polymorphisms and use the LD structure to identify haplotype tagging SNPs, to identify potential functional variants in the promoter region, and to develop tools for association studies of CRP.

Public domain databases with information on common SNPs are now emerging and are widely available. There is a large amount of information not only on the presence of SNPs, but their allele frequencies in different populations and their association with each other (LD structure). The NCBI Single Nucleotide Polymorphism database (dbSNP), the HapMap database, and the Programs for Genomic Application (PGA) SeattleSNPs database were searched in order to construct a consensus SNP map of the CRP gene. A total of 31 polymorphisms were identified from collating data from all the databases in a region of 6050bp that contains the total length of the CRP gene and 5' and 3' flanking sequences. The majority of the polymorphisms are located in the flanking regions, 10 in the 5' region and 14 in the 3' region. Two intronic variants were identified, one a single base substitution and the other a (GT) repeat. Only one coding polymorphism was identified in exon 2, however, this substitution does

not result in any amino acid changes. Additionally, 4 polymorphisms were found located in the 3' untranslated region (UTR).

The LD between the CRP polymorphisms was examined using SNP data from individuals in the SeattleSNPs and HapMap databases using the r^2 LD statistic. SNPs were then grouped in bins according to their LD. Broadly speaking, the total number of bins equates the number of haplotypes. In the African descent population, eighteen bins were generated, and in the European populations, seven bins were generated. The largest bin in the European population contained four polymorphisms, and the largest bin in the African descent population contained six polymorphisms. Although a large number of bins (and therefore haplotypes) were seen in the African descent population, many of the SNPs in single bins had a low minor allele frequency. Therefore, typing less SNPs would be adequate to capture nearly all of the haplotype diversity, achieving a compromise between maximum information on haplotype and minimising the genotyping effort.

Haplotypes based on data from these sets were constructed using the TagIT, PHASE and Tagger haplotype inference programs. Four common haplotypes were identified in a European population with three haplotype tagging SNPs, and eight common haplotypes were identified in the African descent population that capture almost all of the variation in the CRP gene. The results of the TagIT program were concordant with the results from the PHASE and Tagger programs and allowed a tagging set to be selected, based on the location of the haplotype tagging SNPs in relation to other polymorphisms in the gene, and feasibility of genotyping using the TaqMan methodology. This resulted in the selection of the +1444C/T (rs1130864), +2302G/A (rs1205) and +4899T/G (rs3093077) polymorphisms to capture the haplotype diversity in European populations, and the –717A/G (rs2794521), –305G/A (rs3093062), -286C/T/A (rs3091244), +1444C/T (rs1130864), +2302G/A (rs1205) and +4899T/G (rs3093077) polymorphisms to capture the haplotype diversity in African populations.

For the African descent population, seven tagging SNPs were selected using Tagger to generate ten haplotypes. The SeattleSNPs database has genotype information on the –286C/T/A polymorphism, which was selected as part of the tagging SNP set and defines two haplotypes as there are two rare alleles. Since the HapMap database did not have any information on this polymorphism, additional SNPS had to be selected in the Tagger program to generate these haplotypes in the African population. Overall, there was complete consistency in the number of common haplotypes inferred in both European and African descent populations, irrespective of the program used (SeattleSNPs, TagIT or Tagger), or the source data (SeattleSNPs or HapMap).

Indeed, it is now known that tagging SNPs appear portable to other populations of similar ancestry (Need & Goldstein 2006; Conrad *et al.* 2006; de Bakker *et al.* 2006).

5.6 Conclusions

A data mining approach from public domain databases has been used to identify polymorphisms within and flanking the CRP gene. This data has been used to construct a consensus SNP map of the CRP gene containing a total of 31 polymorphisms. By taking advantage of the presence of linkage disequilibrium within the gene, it has been possible to generate haplotype tagging SNPs. This means that the genotyping effort could be reduced from twelve SNPs to just three SNPs in a European population and from thirty SNPs to just six SNPs in an African population, if valid. This panel of haplotype tagging SNPs could therefore be used to conduct genotype-intermediate phenotype and genotype-disease association studies. This work also identified polymorphisms with potential regulatory function, such as the –757T/C, -717A/G and –286C/T/A polymorphisms that are present in both European and African descent populations, and the –756G/A, -305G/A and –94C/T polymorphisms that are not present in European populations.

In Chapter seven, the three tagging SNPs identified in European populations were utilised to examine association between genetic variation in the CRP gene and CRP concentrations and established risk factors for CHD. In Chapters eight and nine, the six tagging SNPs identified in African populations have been utilised to examine association between the CRP gene and CRP concentrations following an inflammatory stimulus, and also to examine the consistency of genotype-CRP associations in populations of differing ancestry.

6. Effect of genotype on CRP concentration; a systematic review and meta-analysis of published data

6.1 Aim

To obtain precise estimates of the effect of common variation in the CRP gene on CRP concentration using data from published studies.

6.2 Background

6.2.1 Genetic association studies

The purpose of genetic association studies is to determine whether there is a causal relationship between the presence of a specific DNA variation (genotype or haplotype) and inherited traits such as cholesterol or blood pressure, or susceptibility to disease such as stroke or myocardial infarction (Kruglyak 2005). One of the limitations of association studies assessing disease susceptibility is that for the most part, the results have been inconsistent and difficult to replicate. This may be because studies in the last decade have been too small to reliably detect odds ratios for disease risk of around 1.2-1.5 when the size of the genetic effect is small (Zondervan & Cardon 2004). For polymorphisms with a frequency of around 30%, large studies of several thousand cases and a similar number of controls are required to detect association reliably. Therefore, the positive associations seen from small studies have a high likelihood of being falsely true (false positive) and therefore the results may not be replicated in larger studies. Similarly, negative associations that are seen may sometimes be falsely negative in smaller studies, as there is inadequate power to detect association, thus leading to inconsistency (Colhoun *et al.* 2003). Therefore, data on a gene is usually a mixture of true positives, false positives, true negatives and false negatives.

One way to overcome this limitation is to use the statistical method of meta-analysis, whereby data from many studies can be pooled and analysed so that a more accurate picture of association can be found (see Chapter 1). This method will also identify any outlying studies with large effect sizes that have arisen by chance and will also allow examination of publication bias. By combining genetic studies, it may be possible to provide stronger evidence of a small effect (Munafo & Flint 2004). An example of this can be seen with the association between APOE4 and Alzheimer's disease, where the allele frequency is 0.15 and the relative risk (RR) of Alzheimer's disease between E4E4 and E3E3 homozygotes is 11.57, found from a meta-analysis of 42 case-control studies (Rubinsztein & Easton 1999). More recent examples include meta-analyses of haemostatic genes such as factor V Leiden and prothrombin, where individual studies have produced conflicting results regarding associations with coronary disease (Ye *et al.*

2006). Meta-analyses of 191 studies found that the per-allele RR for CHD of factor V 1691A and of prothrombin 20210A were 1.17 (95%CI: 1.08-1.28) and 1.31 (95%CI: 1.12-1.52) respectively, suggesting that they were associated with coronary disease.

However, there can be a strong tendency for publication bias for the preferential publication of small studies with positive associations. This can be shown when meta-analyses are conducted by constructing funnel plots and performing statistical analyses such as the Egger test. An example of where the results of meta-analyses have been overturned due to publication bias is the angiotensin converting enzyme (ACE) insertion/deletion (I/D) polymorphism, where meta-analyses of small studies have been contradicted by large RCTs (Borzak & Ridker 1995; Cappelleri *et al.* 1996; Ioannidis *et al.* 1998). The ACE I/D polymorphism is associated with plasma ACE activity and small studies have found it to be associated with an increase in risk of MI. However, large studies have seen little or no association between this polymorphism and increased risk of MI. Later assessment of meta-analyses of the small studies has found evidence of publication bias (Samani *et al.* 1996; Staessen *et al.* 1997). This suggests selective non-publication of studies with negative results. Large RCTs and recent meta-analyses of large studies suggests that the association between the ACE I/D polymorphism and increased risk of MI may not be real (Agerholm-Larsen *et al.* 1997; Agerholm-Larsen *et al.* 2000). Another theme that has emerged from meta-analyses is the publication of earlier studies with extreme results, followed by larger studies that are negative (Ioannidis 2005).

Despite these concerns, there are some examples of meta-analyses providing evidence for a positive genetic effect on disease risk. For example, previous associations between ApoE genotype and CHD risk have been conflicting. This association was recently investigated by pooling data in a meta-analysis from 48 studies (Song *et al.* 2004). The results of this study showed that carriers of the ε4 allele had an odds ratio of 1.42 (95%CI: 1.26-1.61) for CHD risk compared to ε3 homozygous individuals, showing a positive genetic effect on coronary disease risk.

One way of overcoming publication bias is to develop a network of investigators studying the same genes and the same polymorphisms and pool data in a pre-specified manner, and publishing both positive and negative results. In contrast to gene-disease studies, genetic association studies of continuous phenotype traits seem to be more reliable. This is because intermediate phenotypes are continuous traits, so the power of any study is increased compared to studies where the outcome is categorical (e.g. disease event). They are also less prone to misclassification. Also, these intermediate phenotypes are usually the product of the gene itself, and are a more immediate consequence of genetic variation and therefore influenced by a

smaller range of modifying genes and exposures than the disease, so the signal-noise ratio in genotype-intermediate phenotype studies is likely to be much greater.

One important aspect of studying gene-intermediate phenotype associations is that they provide a tool to conduct Mendelian randomisation analyses to investigate the causal relevance of phenotype traits for CHD. For example, a *Taq*IB polymorphism in the cholesteryl ester transfer protein (CETP) gene was robustly associated with differences in HDL-cholesterol levels when the data were pooled in a meta-analysis involving over 10000 individuals, where B2B2 homozygotes had higher HDL levels of 0.12 mmol/L than B1B1 homozygotes (95%CI: 0.11-0.13) (Boekholdt & Thompson 2003). Studies were then conducted to examine the genotype-CVD association, however, these have produced conflicting results. Mendelian randomisation analyses have also been conducted to determine the role of homocysteine in stroke risk. The 677TT genotype in the methyltetrahydrofolate reductase (MTHFR) gene is associated with homocysteine levels that are higher by around 1.7μmol/L compared to individuals homozygous for the C-allele. In a meta-analysis of published studies, TT subjects had an OR for stroke of 1.26 (95%CI: 1.14-1.40) compared to CC subjects, providing evidence of a causal link between homocysteine and stroke (Casas *et al.* 2005).

Polymorphisms have been described in the CRP gene in early, small studies, usually of single SNPs and have provided evidence for an effect on CRP concentration, which, if reliable, would provide a tool to undertake Mendelian randomisation analyses for CRP.

6.2.2 Published data on polymorphisms in CRP

Sequencing of the CRP gene more than a decade ago initially identified a total of nine polymorphisms within the intron, coding regions and 3' untranslated region (UTR). The first polymorphism identified was a multiallelic GT repeat within the intron, which can be repeated between 15-24 times (Goldman *et al.* 1987; Weber *et al.* 1990). Later studies investigated whether this dinucleotide repeat polymorphism was associated with variation in baseline CRP and found significant association (Szalai *et al.* 2002).

Initial studies investigating the effect of the polymorphisms identified from sequencing on CRP concentrations had small sample sizes and tended to focus on the relationship between a single SNP and CRP concentration. Much work was carried out on the +1059G/C non-coding polymorphism in exon 2, which was also one of the first variants to be reported (Cao & Hegele 2000) and subsequently the C-allele was shown to be associated with reduced CRP concentration (Zee & Ridker 2002; Russell *et al.* 2003). Although numerous genotype-CRP association studies have examined the +1059G/C SNP, many of the other polymorphisms identified by the initial sequencing have also been studied, including the +2302G/A variant that

lies downstream of the 3'UTR. This variant was also associated with CRP concentration in a small study of less than 600 patients, where the A-allele was associated with lower CRP concentration compared to the G-allele (Russell *et al.* 2003). This difference in CRP was greater among individuals with both the +1059CC and +2302AA genotypes than those with either one alone. This study was carried out in individuals with systemic lupus erythematosus (SLE) and found that the +2302G/A polymorphism was also associated with SLE.

A small study carried out on another 3' variant, the +1444C/T, found it to be strongly associated with CRP both basally and after an acute inflammatory response in patients following coronary artery bypass surgery (CABG) and intense exercise (Brull *et al.* 2003). Although this polymorphism lies in the 3'UTR, it does not appear to disrupt any known consensus sequences associated with altered mRNA stability (Russell *et al.* 2003).

Following resequencing of the CRP gene, larger studies have been carried out examining association between gene variants and CRP concentration. These studies are more population-based studies carried out in healthy individuals rather than disease groups such as SLE or arterial thrombosis (Zee & Ridker 2002). For example, the +1059G/C association with CRP concentration has been replicated in a number of large studies such as the British Women's Heart and Health study (3529 women) (Davey Smith *et al.* 2005b), the Coronary Artery Risk Development in Young Adults (CARDIA) study (3164 participants) (Carlson *et al.* 2005), and a large community-based study in the United States (2397 participants) (Suk *et al.* 2006).

Resequencing of the gene has identified additional polymorphisms, including a triallelic –286C/T/A variant that lies within a USF1 transcription factor binding site in the promoter. This polymorphism was first studied in patients with SLE, and was found to be in strong LD with the +1444C/T variant in the 3'UTR (D' 0.97) (Russell *et al.* 2003). A later study by Kovacs et al also found strong LD between the two SNPs (D' 0.92) and showed that –286TT individuals had higher CRP levels than CC or CT individuals, and carriage of the A-allele resulted in even higher CRP concentrations that those seen in TT individuals (Kovacs *et al.* 2005).

Another polymorphism, the +194A/T SNP, was also identified from gene resequencing and has also been found to be in almost complete LD with the +1444C/T variant in a recent publication examining CRP polymorphisms in the Womens Health study (WHS), the Physicians Health study (PHS) and the Pravastatin Inflammation/ CRP Evaluation (PRINCE) study (r^2 0.911-0.982) (Miller *et al.* 2005). As expected, the T-allele of the +194A/T SNP is also in very strong LD with the T-allele of the –286C/T/A variant, and lies within the same haplotype block as the +1444C/T SNP in Caucasians. This polymorphism lies 29bp downstream of exon 1,

within the intron close to the $(GT)_n$ polymorphic site and was initially chosen to be studied due to its close proximity to the exonic +1059G/C SNP (Zee & Ridker 2002).

Recent association studies have turned their attention to examining more than one SNP and have either reported the effects of several polymorphisms on CRP concentration (Miller *et al.* 2005), or have combined SNPs to generate haplotypes and have analysed the effect of haplotype on CRP (Carlson *et al.* 2005). In such cases where more than one polymorphism has been studied, the LD between the variants has also been calculated. This helps to integrate data from different polymorphisms as inferences can be made about how a polymorphism would affect CRP concentration based on LD with other polymorphisms previously studied.

The observation that circulating CRP concentration is subject to genetic regulation has important implications for understanding the role of CRP in the pathogenesis of coronary artery disease and its potential role in coronary risk prediction. The aim of this study was to conduct a systematic review of the literature on the CRP genotype-CRP concentration association to assess whether the effect is real, to assess how large the effect is, and to evaluate whether other covariables relevant to cardiovascular disease are distributed evenly among the genotypic classes. In addition, information on linkage disequilibrium between SNPs was sought to assess to what extent information on single SNPs could be combined.

6.3 Methods

6.3.1 Data search

Three electronic databases (PubMed Medline, EMBASE and the Cochrane Collaboration library) were searched up to and including January 2006 for all studies evaluating the association between CRP polymorphisms and CRP levels. For the search, the text words, which were also MeSH terms, "polymorphism", "mutation", "genes", "genetic", "variant" and "SNP" in combination with "c-reactive protein" and "CRP" were used. The literature search was limited to "human", and "English Language". Abstracts from major cardiovascular related conferences in the last 5 years were identified. Any additional studies in the references of all identified publications were also searched.

6.3.2 Selection criteria

For inclusion, studies had to have an analytical design (case-control, prospective or cross-sectional) and examine the association between polymorphisms in the CRP gene and unadjusted basal CRP concentration in healthy individuals. In all searches, when relevant information was not reported, the authors were contacted at least twice to obtain the required information, using a standard tabular format for data.

6.3.3 Data extraction

Data for analysis including study design, genotyping methods, CRP measurement methods, mean age of participants, gender, frequency of genotypes and alleles, unadjusted CRP concentrations, ethnic background, and cardiovascular risk factors were extracted and entered into an Excel spreadsheet. The analysis was limited to the controls from case-control studies because disease might modify the effect of the gene on CRP concentrations. Where more than one polymorphism had been studied, information about the LD between them was noted.

6.3.4 Statistical analysis

Genotype frequencies were compared to those expected under the assumption of Hardy-Weinberg equilibrium by χ^2 analysis. Meta-analyses were conducted for each polymorphism to obtain the weighted mean difference (WMD) in plasma CRP levels between individuals of differing genotype. For these analyses a random effects model was used, in order to allow for any heterogeneity across studies. The geometric mean was used wherever possible, or calculated from the arithmetic mean. However, some studies only used median CRP values and it was not possible to obtain geometric mean values from these authors. In some cases, the CRP values had to be calculated from a graph, as a tabular format was not available. The majority of studies reported standard deviation (SD) values, although where these values were not available, the SD from the largest study was used. In cases where the SD values were implausible due to small studies sizes, the SD from the largest study was also used. In addition, "per allele" effects were determined using regression analysis. The DerSimonian and Laird Q test, and the I^2 test (Higgins *et al.* 2003) were used to evaluate the degree of heterogeneity between studies, and funnel plots to evaluate small-study bias, of which publication bias is one potential cause. In addition, the Egger test was also used to test for asymmetry in the funnel plots and the presence of bias. Data were analysed using the Comprehensive Meta-analysis (CMA) version 2.0 software (Biostat) and Stata 8.0 (Stata Corporation, College Station, Texas, 2003).

6.4 Results

6.4.1 Studies obtained for meta-analysis

Authors from 27 publications (24 full length articles and 3 abstracts) were contacted for data on CRP polymorphisms and CRP concentrations and other measured variables. Out of these, 10 authors did not respond, even when contacted more than twice. Four authors declined to send data although in some cases, much of the data could be extracted from the publication. The rest of the authors sent data either in a tabular format or the whole database.

A total of 12 publications (14 studies) were included for data extraction and analysis, and provided information on 7 polymorphisms. The data were entered into an Excel spreadsheet separately according to ethnicity where possible. Data from all 12 publications were used to assess the genotype-CRP concentration association. In addition, 8 publications contained some measures for CHD risk factors such as blood pressure, BMI, LDL-cholesterol, HDL-cholesterol and smoking, allowing evaluation of the genotype-intermediate phenotype association. Where publications had data on more than one study, the data were separated and entered as different studies. Data on genotype were included if two or more studies had investigated that polymorphism.

Overall, 6 polymorphisms were evaluated, the –717A/G polymorphism (rs2794521) (5 publications), -286C/T/A polymorphism (rs3091244) (3 publications), +194A/T polymorphism (rs1417938) (3 publications), +1059G/C polymorphism (rs1800947) (8 publications), +1444C/T polymorphism (rs1130864) (7 publications) and the +2302G/A polymorphism (rs1205) (4 publications). It was not possible to study the –757T/C polymorphism (rs3093059) as it had only been studied in one publication.

Figure 6.1. Flow diagram of the search results for the systematic review examining the association between CRP genotype and CRP concentration.

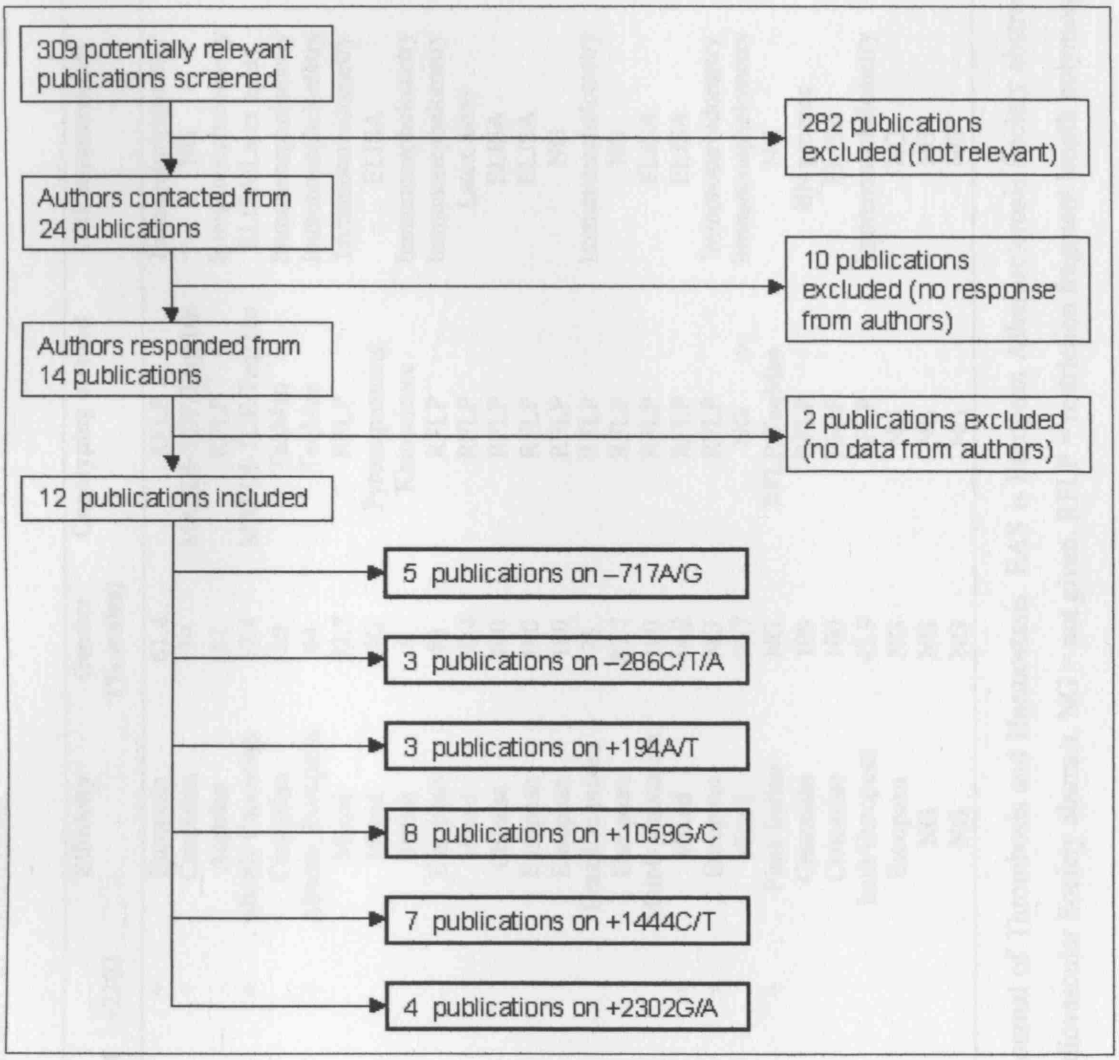


Table 6.1. Characteristics of published studies on CRP polymorphisms and CRP concentration.

Publication	Polymorphism studied						Ethnicity	Gender (% males)	Genotyping method	CRP measurement	Total # subjects
	-717	-286	+194	+1059	+1444	+2302					
Willot 2006	+				+	+	European	61.4	RFLP	Immunoturbidometry	318
Suk Danik 2006	+	+	+	+	+	+	Caucasian	79	MALDI-TOF/TaqMan	NG	1847
Vormittag 2005					+		Austrian	51	RFLP	Immunonephelometry	122
Miller 2005	+	+	+	+	+	+	Mainly Caucasian	57.4	MALDI-TOF/TaqMan	ELISA/Latex assay	3107
Eklund 2005				+			Caucasian	50	TaqMan	Immunonephelometry	335
Carlson 2005	+	+	+	+	+	+	African, European	44	TaqMan	Immunonephelometry	3164
D'Aiuto 2005					+		Mixed	52.7	RFLP	Immunoturbidometry	55
Szalai 2005		+					Mixed	NG	Pyrosequencing	ELISA	309
Davey Smith 2005				+			British	0	Kbioscience	Immunonephelometry	3529
Kovacs 2005	+	+		+	+		European	83	RFLP	Immunonephelometry	579
Suk 2005			+	+			Mixed	70.4	RFLP	Latex assay	2395
Chen 2005	+						Chinese	100	RFLP	ELISA	1234
de Maat 2004				+			European	100	RFLP	ELISA	584
Flex 2004				+			European	100	RFLP	NG	471
Araujo 2004				+			Brazilian mixed	75	RFLP	Immunonephelometry	684
Zee 2004 (Athero)			+	+			European	87.7	RFLP	NG	779
Zee 2004 (J Thr Haem)			+	+			Mainly Caucasian	100	RFLP	ELISA	260
Obisesan 2004	+		+	+		+	Mixed	44.2	RFLP	ELISA	63
Russell 2004		+		+	+	+	European	NG	RFLP	Immunoturbidometry	586
Judson 2004				+			Mixed	46.7	NG	Immunonephelometry	674
Wolford 2003	+		+		+	+	Pima Indian	NG	RFLP/TaqMan	NG	1300
Brull 2003	+			+	+		Caucasian	100	RFLP	BN prospec	443
Zee 2002				+			Caucasian	100	RFLP	ELISA	1452
Hegele 2001				+			Inuit/European	43.9	RFLP	Immunonephelometry	237
Eklund 2005 (EAS)	+			+	+		European	NG	NG	NG	600
Suk 2004 (AHA)	+		+	+	+		NG	NG	NG	NG	2103
Lee 2002 (CCS)				+			NG	NG	NG	NG	1014

Athero is the Atherosclerosis journal and J Thr Haem is the Journal of Thrombosis and Haemostasis. EAS is European Atherosclerosis Society abstract, AHA is American Heart Association abstract and CCS is Canadian Cardiovascular Society abstract. NG = not given. RFLP = restriction fragment length polymorphism.

Table 6.2. Study measures from published data on CRP polymorphisms.

Publication	Other measures reported											
	Age	Gender	BMI	TC	LDL	HDL	Tg	BP	Diabetes	Smoking	Alcohol	Inflammatory
Willot 2006												
Suk Danik 2006												
Vormittag 2005												
Miller 2005												
Eklund 2005		Yes										
Carlson 2005	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
D'Aiuto 2005	Yes	Yes	Yes					Yes		Yes		IL-6
Szalai 2005	Yes	Yes								Yes		
Davey Smith 2005	Yes		Yes			Yes	Yes	Yes	Yes	Yes	Yes	
Kovacs 2005	Yes	Yes	Yes					Yes	Yes	Yes		
Suk 2005	Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes		
Chen 2005	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
de Maat 2004	Yes	Yes	Yes							Yes		IL-6, fibrinogen, TNFα, sICAM-1
Flex 2004	Yes	Yes						Yes	Yes	Yes		
Araujo 2004	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes		
Zee 2004 (Athero)	Yes	Yes	Yes						Yes	Yes		
Zee 2004 (J Thr Haem)	Yes		Yes					Yes	Yes	Yes		
Obisesan 2004	Yes	Yes	Yes									
Russell 2004												
Judson 2004	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Wolford 2003									Yes			
Brull 2003	Yes	Yes	Yes					Yes		Yes		
Zee 2002	Yes		Yes					Yes	Yes	Yes		
Hegele 2001												
Eklund 2005 (EAS)												
Suk 2004 (AHA)												
Lee 2002 (CCS)	Yes	Yes	Yes				Yes					IL-6

Athero is the Atherosclerosis journal and J Thr Haem is the Journal of Thrombosis and Haemostasis. EAS is European Atherosclerosis Society abstract, AHA is American Heart Association abstract and CCS is Canadian Cardiovascular Society abstract. TC = total cholesterol; Tg = triglycerides.

6.4.2 LD between CRP polymorphisms

A total of six publications were identified where more than one polymorphism in the CRP gene had been studied and contained information about the LD between polymorphisms. Some publications gave only the D' measures of LD and some gave both the D' measure and the r^2 correlation coefficient. A total of 7 variants were examined, including the triallelic –286C/T/A SNP, which was categorised as two different SNPs (-286C>T and –286C>A) by Miller et al. to allow more accurate LD values to be determined (Miller *et al.* 2005).

When solely D' measures of LD were considered, all polymorphisms were in high LD (D'=0.61-1.00), however, when LD was examined using solely the r^2 correlation coefficient, high LD was only seen between the T-alleles of the –286C/T/A, the +194A/T and the +1444C/T polymorphisms (r^2=0.97-0.984; D'=0.92-1.00). High LD was also seen between the minor C-allele of the –757T/C polymorphism and the A-allele of the –286C/T/A polymorphism (r^2=0.975-1.00; D'=1.00). The D' measure is less stringent than the r^2 measure in that it is less sensitive to differences in minor allele frequency whereas r^2 can only have a value of 1.0 when the polymorphisms have the same allele frequency, which may account for the differences in LD seen (see Table 6.3).

Table 6.3. LD between CRP polymorphisms from published data.

	-757 T>C	-717 A>G	-286 C>T	-286 C>A	+194 A>T	+1059 G>C	+1444 C>T	+2302 G>A
-757 T>C		1.000 1.000 0.914	1.000 1.000 0.835	1.000 1.000 1.000	1.000 1.000 0.846	1.000 0.992 1.000	1.000 0.946 1.000	0.610 1.000 0.911
-717 A>G	0.026 0.027 0.024		1.000 0.994 0.957 -0.97	1.000 1.000 0.912	1.000 0.957 0.955	1.000 1.000 1.000 -1.00	0.984 0.994 1.000 -0.99 0.86	0.848 0.994 0.985 0.86
-286 C>T	0.026 0.032 0.024	0.154 0.176 0.160		n/a n/a n/a	1.000 0.975 1.000	1.000 0.957 1.000 0.91 -0.89	0.989 0.985 0.963 0.97 0.92	0.766 0.994 0.988 0.98
-286 C>A	1.000 1.000 0.975 ~0.99	0.027 0.027 0.024	n/a n/a n/a		1.000 1.000 1.000	1.000 0.933 1.000	0.989 0.948 1.000	0.766 1.000 1.000
+194 A>T	0.025 0.032 0.025	0.149 0.165 0.156	0.984 0.945 0.982	0.026 0.032 0.035		1.000 1.000 1.000	1.000 0.962 0.960	0.786 0.949 1.000
+1059 G>C	0.005 0.005 0.004	0.035 0.028 0.026	0.034 0.030 0.030	0.006 0.004 0.005	0.033 0.033 0.029		1.000 0.956 1.000 0.91 -0.73 0.98	0.758 0.979 0.919 0.96 0.98
+1444 C>T	0.025 0.028 0.032	0.150 0.176 0.174	0.974 0.966 0.907	0.026 0.028 0.033	0.982 0.917 0.911	0.034 0.029 0.030 0.03		0.746 0.972 0.976 0.96 0.98 0.97
+2302 G>A	0.015 0.034 0.029	0.166 0.189 0.181	0.130 0.216 0.218	0.016 0.034 0.035	0.132 0.199 0.219	0.086 0.141 0.110 0.14	0.125 0.207 0.212 0.2	

D' LD is on the upper right and r^2 LD is on the lower left of the table.

Key:
Miller *et al.* 2005 – WHS (n=717)
Miller *et al.* 2005 – PRINCE (n=1437)
Miller *et al.* 2005 – PHS controls (n=696)
Suk Danik *et al.* 2006 (n=1847)
Russell *et al.* 2003 (n=586)
Kovacs *et al.* 2004 (n=579)
Timpson *et al.* 2005 (n=3218)
Willot *et al.* 2006 – Crohn's Disease controls (n=206)

6.4.3 –717A/G polymorphism and CRP concentration

Data from 5 studies of Caucasian background totalling 4765 subjects were used for this analysis (Willot *et al.* 2006a; Suk Danik *et al.* 2006; Miller *et al.* 2005; Kovacs *et al.* 2004). The majority of subjects in each study were male (>61%), however, it was not possible to separate data according to gender. The frequencies of the A- and G- alleles were assessed by χ^2 analysis and all the studies were in Hardy-Weinberg equilibrium. The CRP concentrations were compared by meta-analysis for all the genotypes (AA vs. AG, AA vs. GG and AG vs. GG) and the summary data were obtained and forest plots were constructed (see Figure 6.2).

Figure 6.2a. Forest plot showing the weighted mean difference in CRP concentration between the -717AA vs. –717AG genotypes using a random effects model.

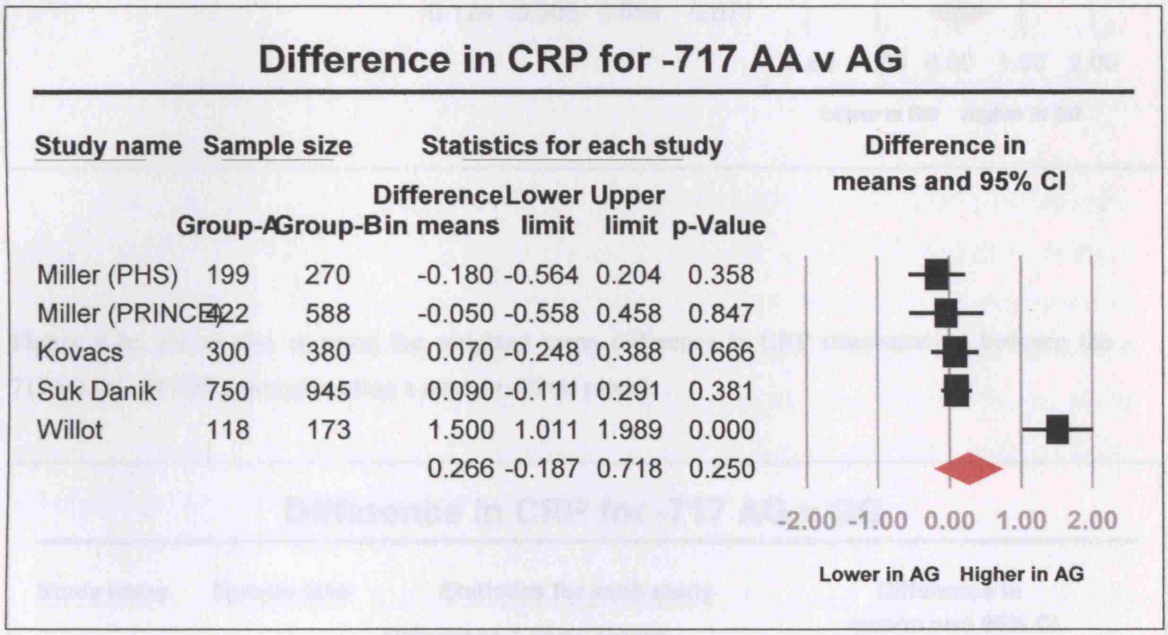


Figure 6.2b. Forest plot showing the weighted mean difference in CRP concentration between the -717AA vs. –717GG genotypes using a random effects model.

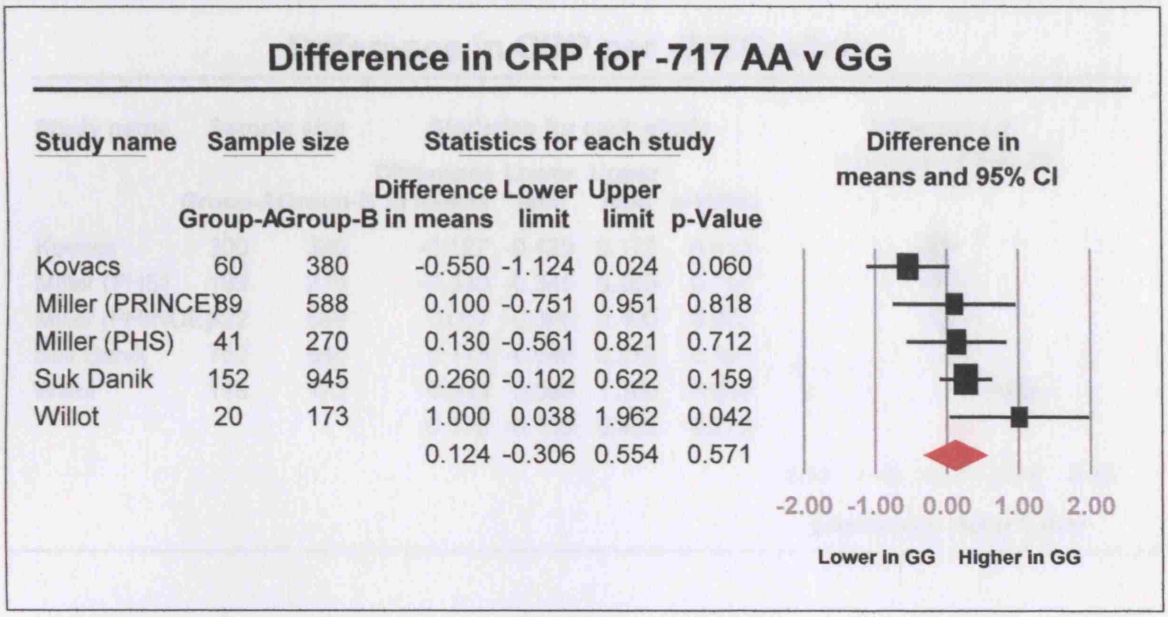


Figure 6.2c. Forest plot showing the weighted mean difference in CRP concentration between the -717AG vs. –717GG genotypes using a random effects model.

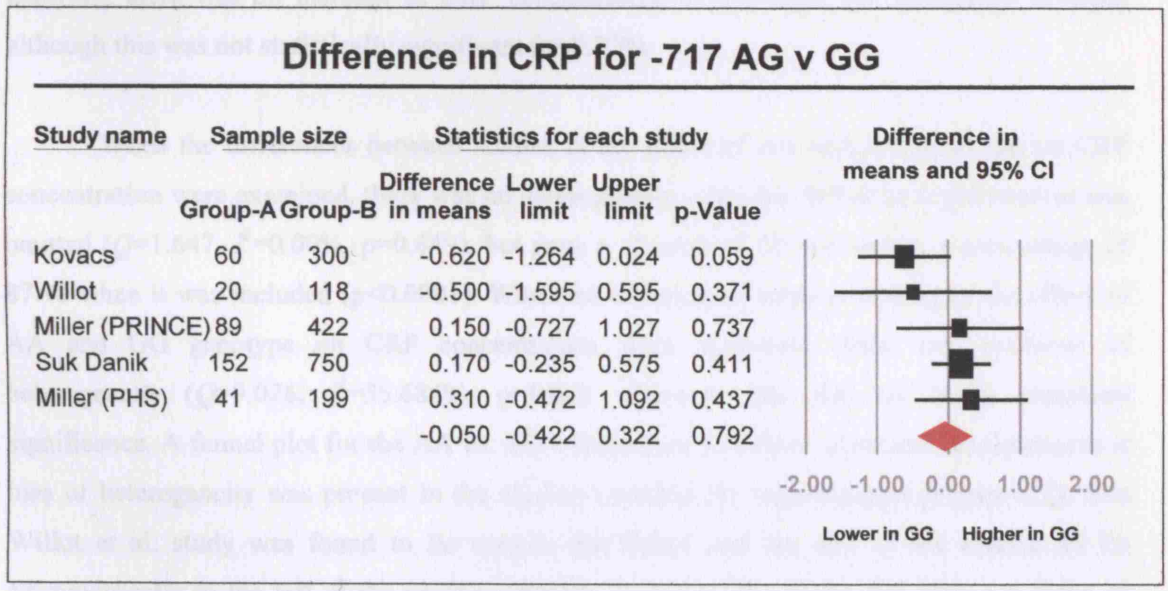
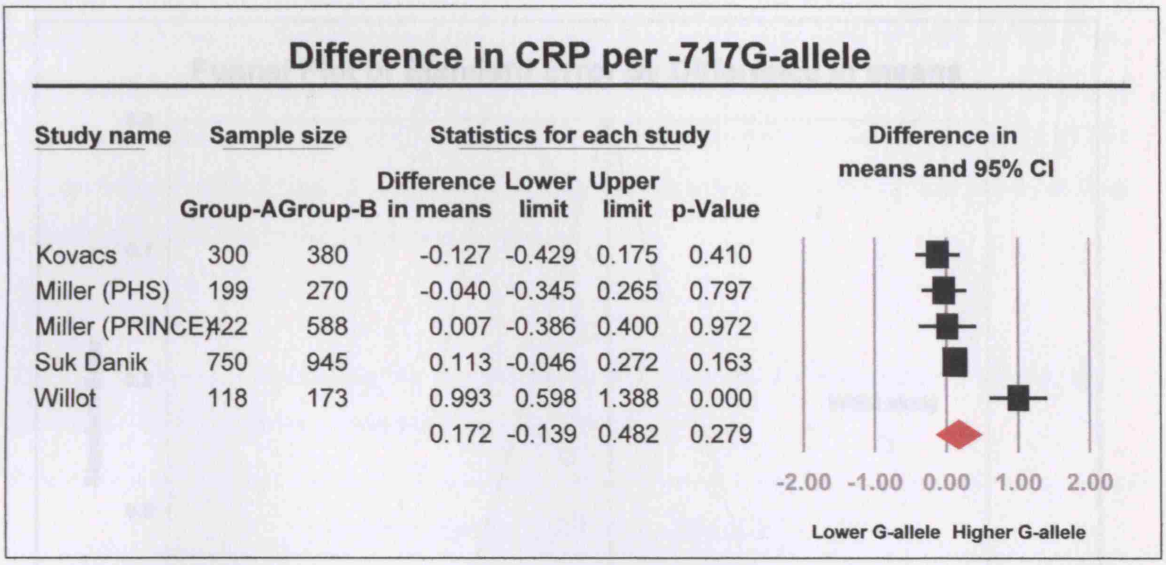


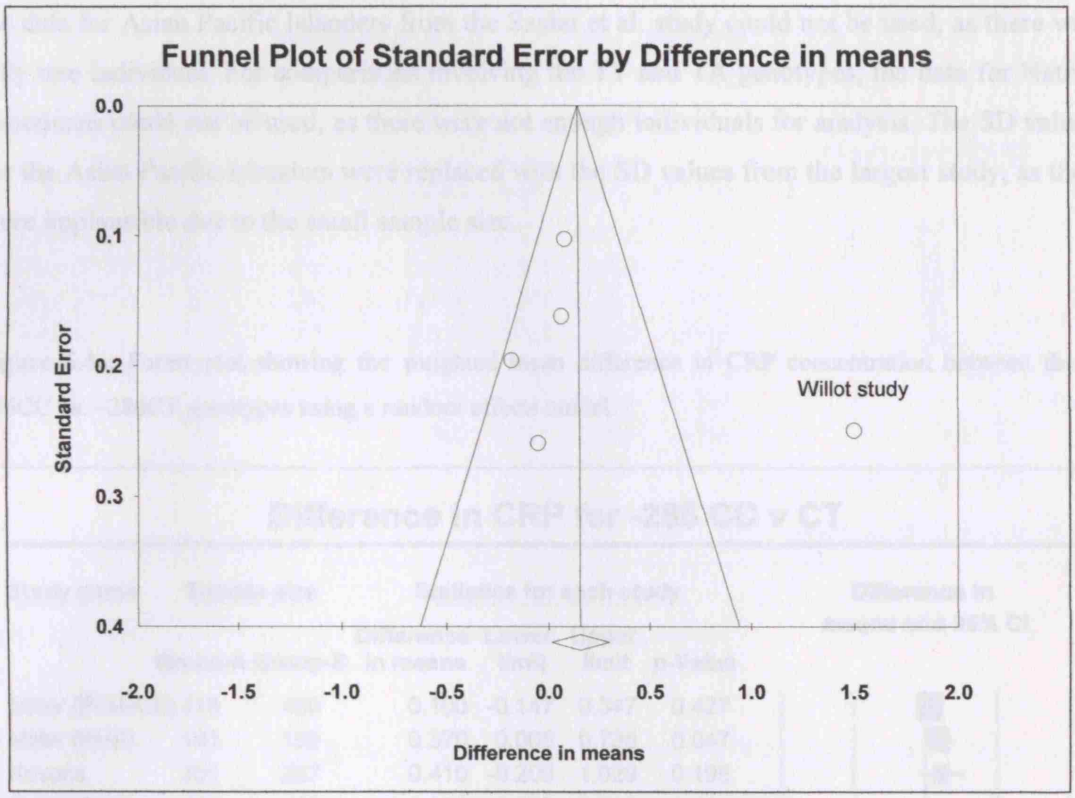
Figure 6.2d. Forest plot showing the weighted mean difference in CRP concentration in per allele analyses.



These results suggest that there is no significant difference in CRP concentration between individuals with a –717AA genotype and those with a –717GG genotype using a random effects model. When individual studies were examined, no significant difference in CRP between genotypes was reported, except for the Willot et al. study, which showed a significant difference between AA and AG individuals (1.5mg/L, p<0.00001). In per allele analyses, there was an increase in CRP concentration of 0.17mg/L per additional G-allele, although this was not statistically significant (p=0.279).

When the differences between studies in the effect of AA and AG genotype on CRP concentration were examined, there was no heterogeneity when the Willot et al publication was omitted (Q=1.647, I^2=0.00%, p=0.649), but gave a Q value of 33.269 and an I^2 percentage of 87.98 when it was included (p<0.0001). When the differences between studies in the effect of AA and GG genotype on CRP concentration were examined, there was evidence of heterogeneity (Q=9.026, I^2=55.684%, p=0.06), although this did not reach statistical significance. A funnel plot for the AA vs. AG comparison was then constructed to determine if bias or heterogeneity was present in the studies included for meta-analysis (Figure 6.3). The Willot et al. study was found to lie outside the funnel and the rest of the studies all lie asymmetrically to the left of the pooled estimate. However, the Egger test gave a p value of 0.539, which was not significant for presence of bias. When baseline characteristics of the Willot et al. study were compared to those from the other studies, attributes such as ethnicity, age, and BMI appeared to be similar, although this study had a smaller sample size and therefore the more extreme difference in CRP concentration by genotype may be due to chance.

Figure 6.3. Funnel plot to show presence of bias in the studies used in the meta-analysis to compare CRP concentration between -717AA and -717AG individuals.



The effect of this polymorphism on other intermediate phenotypes could not be examined, as other measures apart from CRP were only available in the study by Kovacs et al., although no association was seen between genotype and age or BMI.

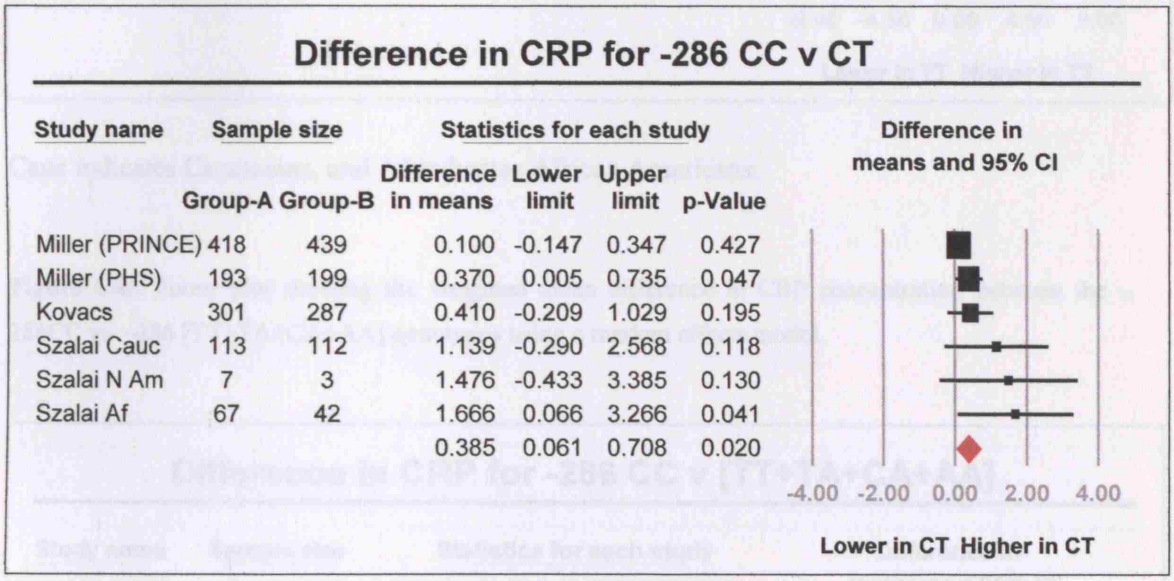
6.4.4 –286C/T/A polymorphism and CRP concentration

Data from 4 studies involving a total of 3223 subjects were used for this analysis (Miller *et al.* 2005; Szalai *et al.* 2005; Kovacs *et al.* 2004). For the study by Szalai et al., the data were separated according to ethnicity, which included 280 Caucasians, 207 African Americans, 16 native Americans and 8 Asian Pacific Islanders. The majority of subjects in the other studies were male (>72%), however, it was not possible to ascertain the percentage of male subjects in the Szalai et al. publication. The frequencies of the alleles were assessed by χ^2 analysis and were in Hardy-Weinberg equilibrium.

The CRP concentrations by genotype were compared by meta-analysis and the summary data were obtained and forest plots were constructed (see Figure 6.4). Additionally,

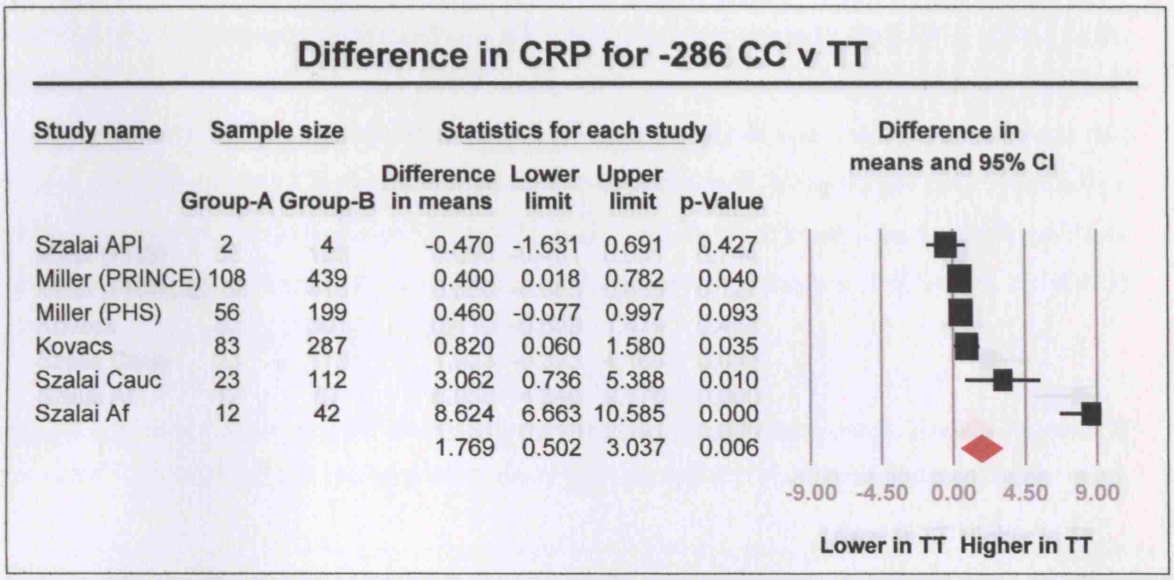
the mean CRP data for the TT, TA, CA and AA genotypes were combined and compared with the CC and CT genotype data, as the A-allele has a low frequency in Europeans so there were fewer genotypes containing the A-allele. For comparisons involving the CT or CA genotypes, the data for Asian Pacific Islanders from the Szalai et al. study could not be used, as there was only one individual. For comparisons involving the TT and TA genotypes, the data for Native Americans could not be used, as there were not enough individuals for analysis. The SD values for the Asian Pacific Islanders were replaced with the SD values from the largest study, as they were implausible due to the small sample size.

Figure 6.4a. Forest plot showing the weighted mean difference in CRP concentration between the –286CC vs. –286CT genotypes using a random effects model.



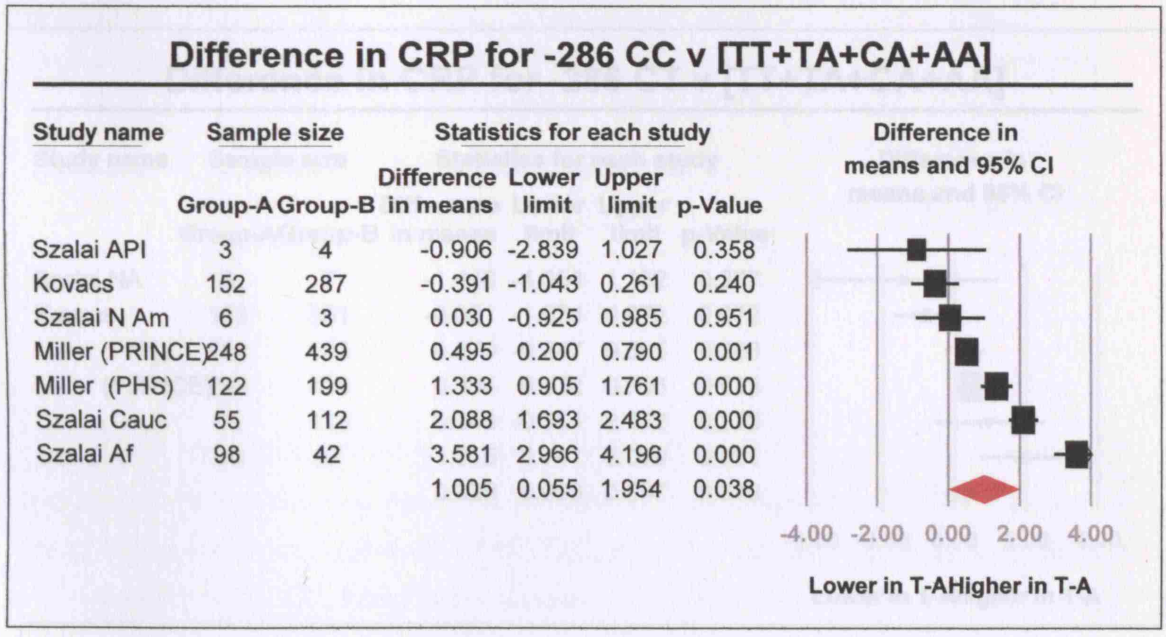
N Am indicates Native Americans, Cauc indicates Caucasians, and Af indicates African Americans.

Figure 6.4b. Forest plot showing the weighted mean difference in CRP concentration between the –286CC vs. –286TT genotypes using a random effects model.



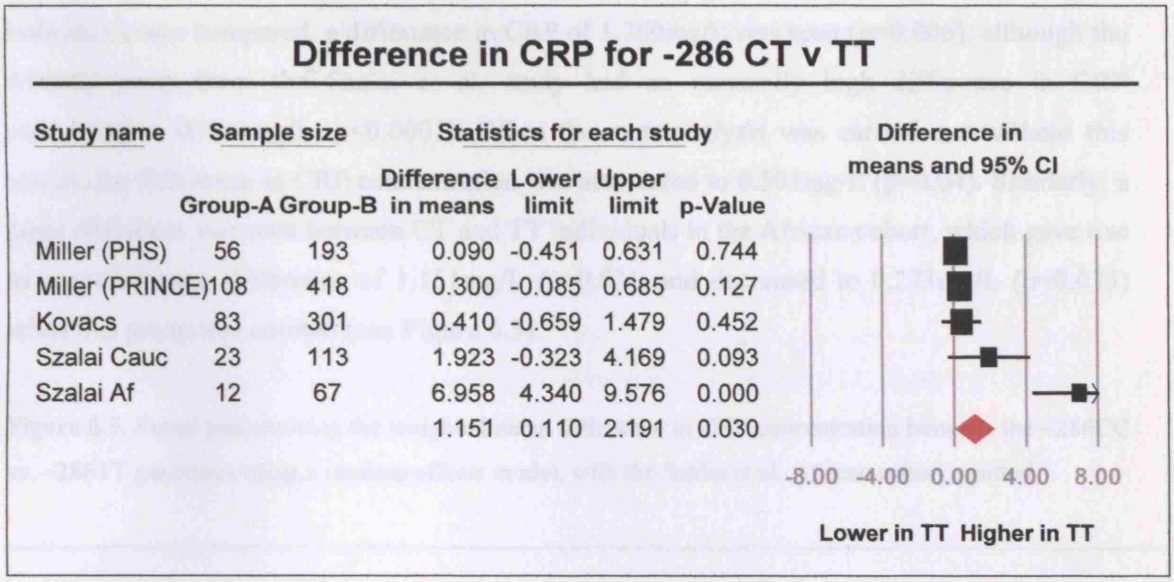
Cauc indicates Caucasians, and Af indicates African Americans.

Figure 6.4c. Forest plot showing the weighted mean difference in CRP concentration between the –286CC vs. –286 [TT+TA+CA+AA] genotypes using a random effects model.



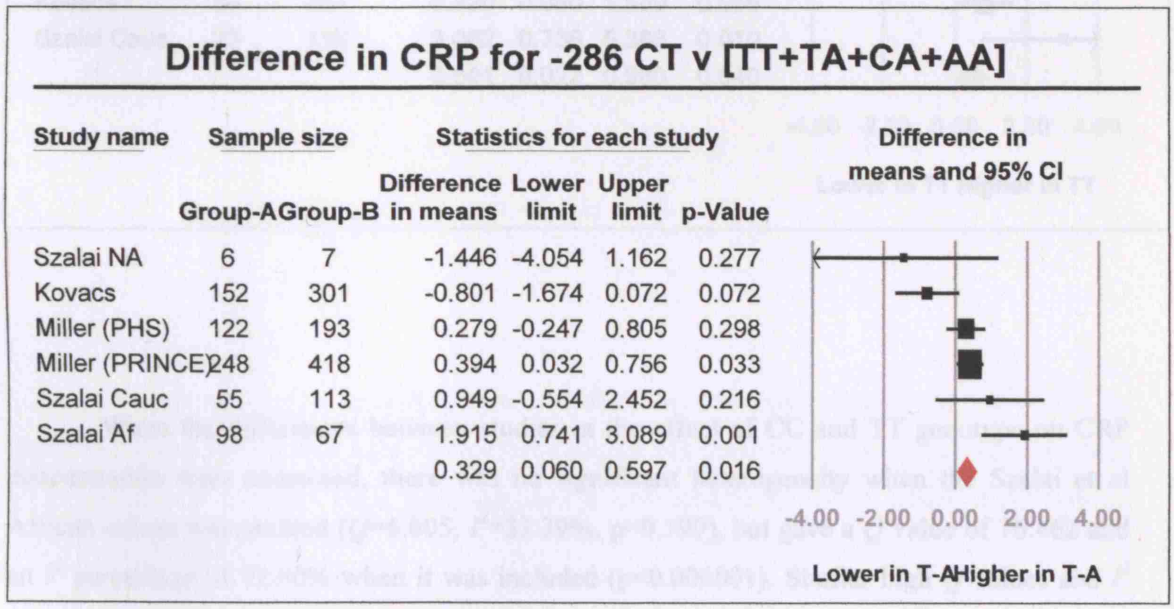
API indicates Asian Pacific Islanders, N Am indicates Native Americans, Cauc indicates Caucasians, and Af indicates African Americans.

Figure 6.4d. Forest plot showing the weighted mean difference in CRP concentration between the –286CT vs. –286TT genotypes using a random effects model.



Cauc indicates Caucasians, and Af indicates African Americans.

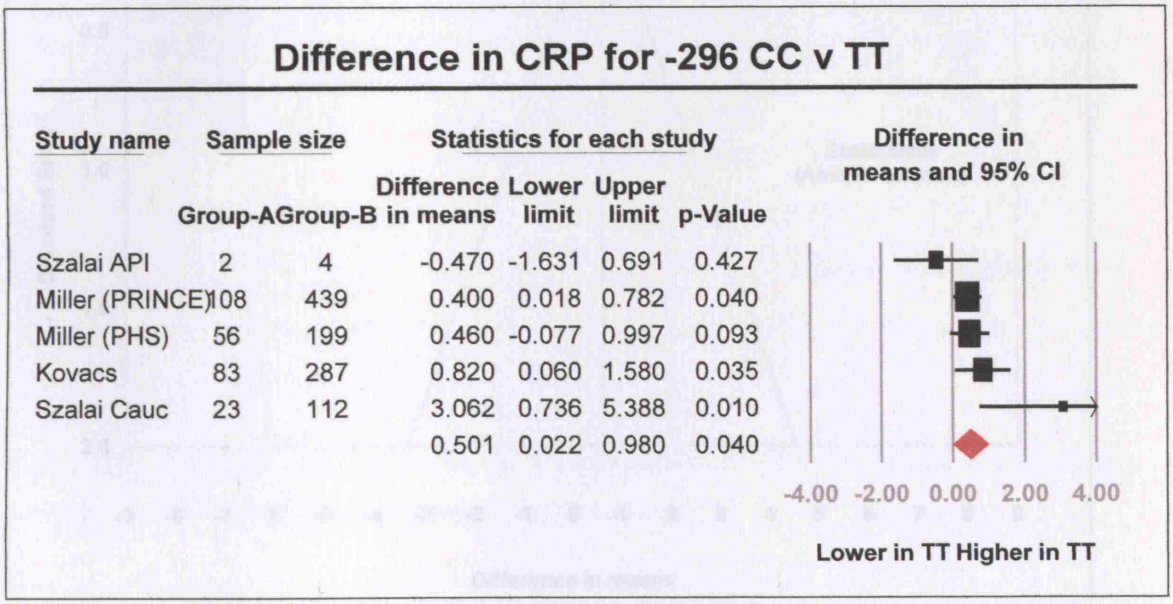
Figure 6.4e. Forest plot showing the weighted mean difference in CRP concentration between the –286CT vs. –286 [TT+TA+CA+AA] genotypes using a random effects model.



N Am indicates Native Americans, Cauc indicates Caucasians, and Af indicates African Americans.

These results suggest that there is a small significant difference in CRP concentration of 0.385mg/L between individuals with a –286CC genotype and those with a –286CT genotype (p=0.02) using a random effects model. When the mean CRP concentration between CC and TT individuals was compared, a difference in CRP of 1.769mg/L was seen (p=0.006), although the African group from the Szalai et al. study had an unusually high difference in CRP concentration (8.624mg/L; p<0.0001). When the meta-analysis was carried out without this cohort, the difference in CRP concentration was attenuated to 0.501mg/L (p=0.04). Similarly, a large difference was seen between CT and TT individuals in the African cohort, which gave rise to a pooled mean difference of 1.151mg/L (p=0.03), and decreased to 0.273mg/L (p=0.073) when this group was omitted (see Figure 6.5).

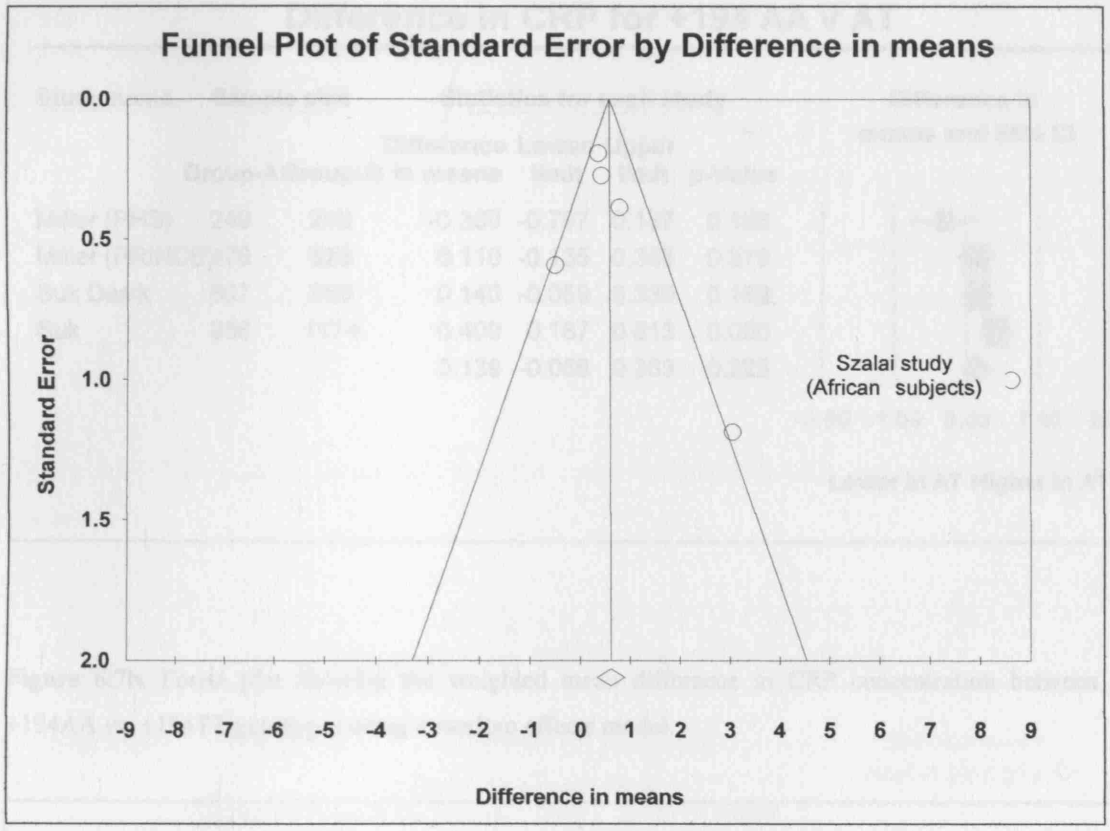
Figure 6.5. Forest plot showing the weighted mean difference in CRP concentration between the –286CC vs. –286TT genotypes using a random effects model, with the Szalai et al. African cohort omitted.



When the differences between studies in the effect of CC and TT genotype on CRP concentration were examined, there was no significant heterogeneity when the Szalai et al African cohort was omitted (Q=6.005, I^2=33.39%, p=0.199), but gave a Q value of 70.462 and an I^2 percentage of 92.90% when it was included (p<0.000001). Similar high Q values and I^2 percentages were seen in the comparisons between CT and TT genotypes, CC and [TT+TA+CA+AA] genotypes and CT and [TT+TA+CA+AA] genotypes, which were substantially reduced when the Szalai et al. African cohort was omitted. A funnel plot for the CC vs. TT comparison was then constructed to determine if bias or heterogeneity was present in the studies included for meta-analysis (see Figure 6.6). The African cohort from the Szalai et al.

study was found to lie outside the funnel and the rest of the studies were mainly clustered towards the top of the funnel. The Egger test gave a p value of 0.172, which was not significant for presence of bias. When all the studies were compared, the baseline characteristics appeared similar across studies and differences were seen only for sample size and ethnicity, which may account for the heterogeneity seen.

Figure 6.6. Funnel plot to show if bias is present in the meta-analysis to compare CRP concentration between –286CC and –286TT individuals.



The effect of this polymorphism on other intermediate phenotypes could not be examined, as other measures apart from CRP were not available except in the Szalai et al. study where there was no association with age, and the Kovacs et al. study, where there was no association with age or BMI.

6.4.5 +194A/T polymorphism and CRP concentration

Data from 4 studies totalling 6375 individuals were used for this analysis (Suk Danik *et al.* 2006; Miller *et al.* 2005; Suk *et al.* 2006). The majority of subjects were male (>72%), and were all Caucasian except for the subjects in the Suk et al. study, where 19% were of different

ethnicity, although it was not possible to separate the data according to ethnicity. The frequencies of the alleles were assessed by χ^2 analysis and were in Hardy-Weinberg equilibrium. The CRP concentrations were compared by meta-analysis for all the genotypes (AA vs. AT, AA vs. TT and AT vs. TT) and the summary data were obtained and forest plots were constructed (see Figure 6.7).

Figure 6.7a. Forest plot showing the weighted mean difference in CRP concentration between the +194AA vs. +194AT genotypes using a random effects model.

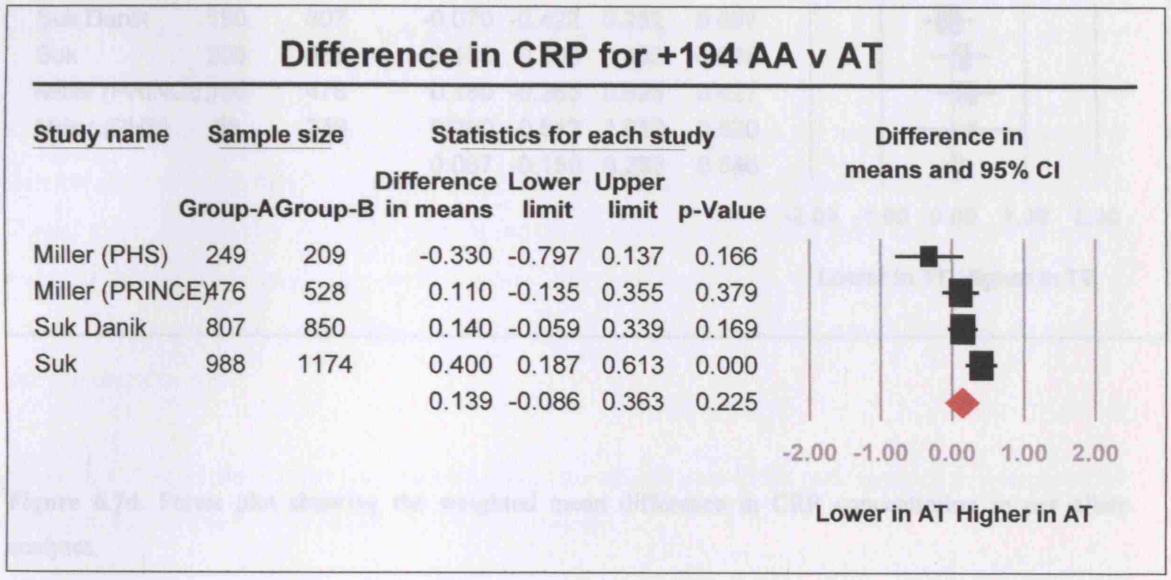


Figure 6.7b. Forest plot showing the weighted mean difference in CRP concentration between the +194AA vs. +194TT genotypes using a random effects model.

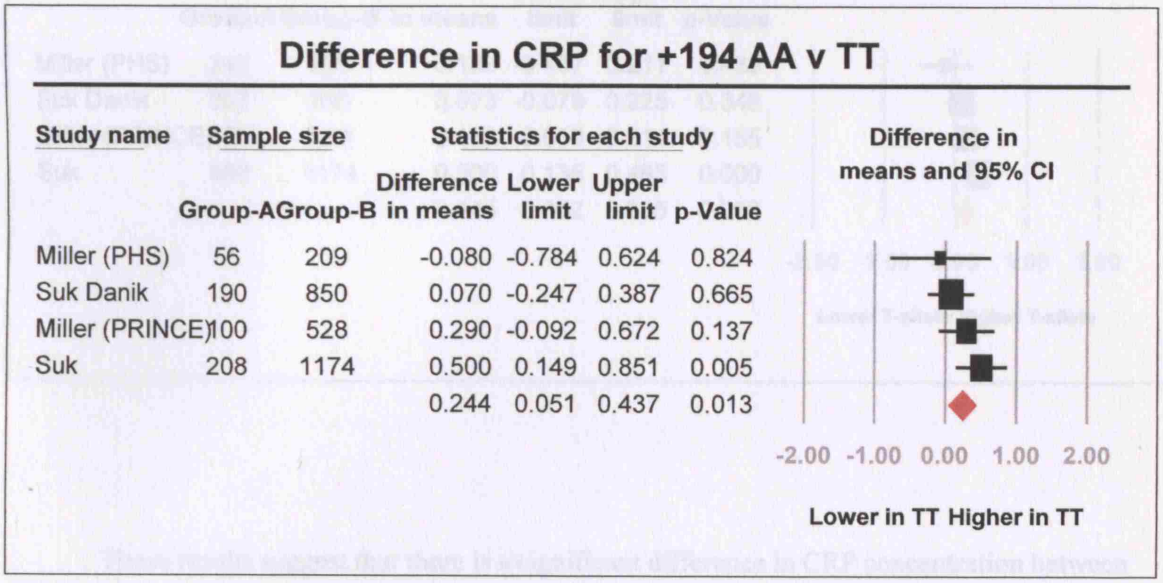


Figure 6.7c. Forest plot showing the weighted mean difference in CRP concentration between the +194AT vs. +194TT genotypes using a random effects model.

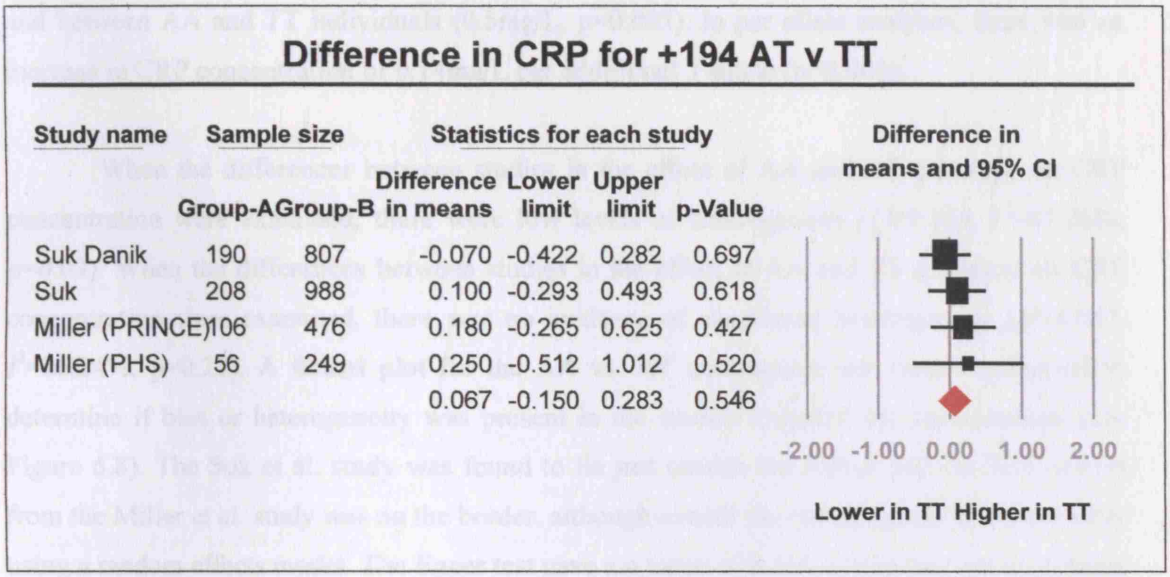
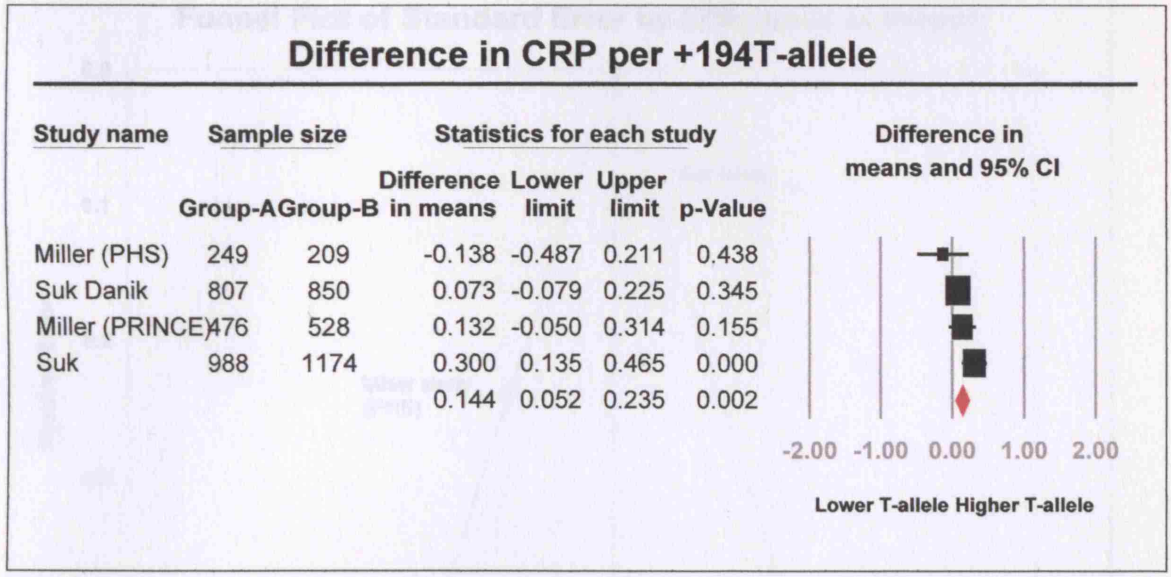


Figure 6.7d. Forest plot showing the weighted mean difference in CRP concentration in per allele analyses.

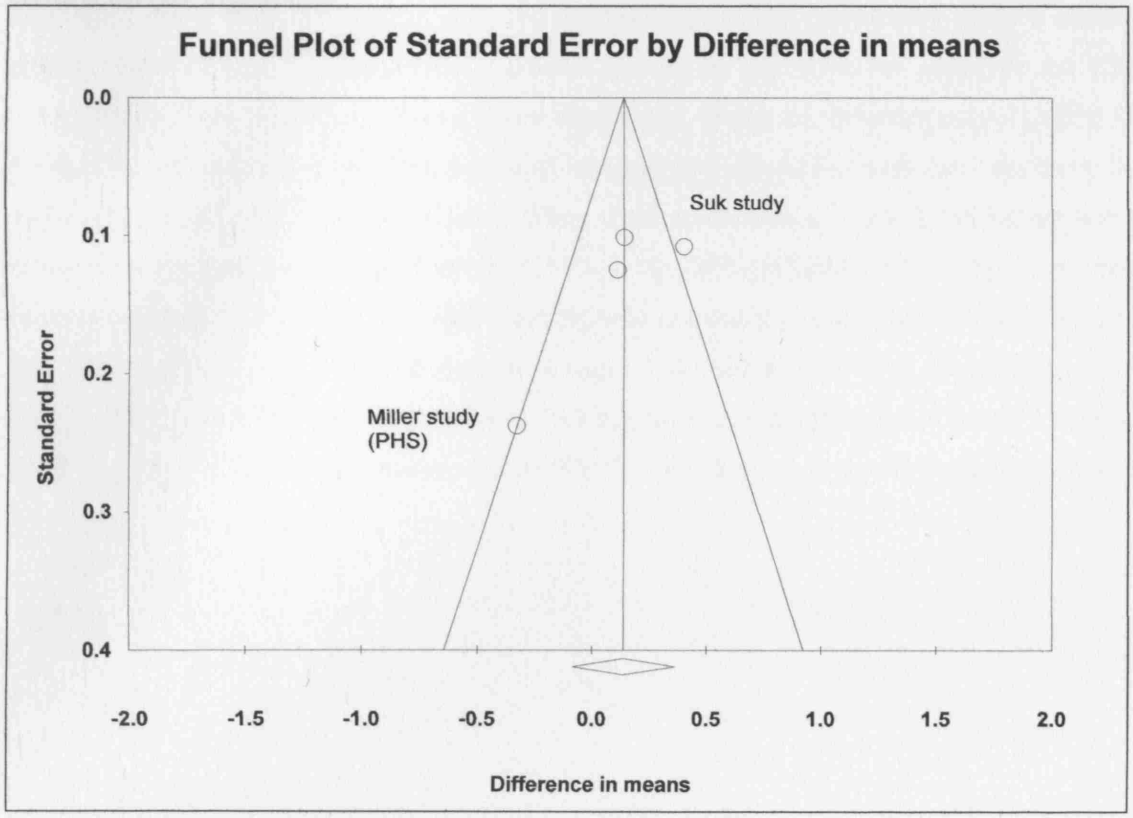


These results suggest that there is a significant difference in CRP concentration between carrying the +194AA genotype and the +194TT genotype with a recessive model of inheritance

(p=0.013). However, no difference was seen between AA and AT genotypes, or between AT and TT genotypes using a random effects model. When individual studies were examined, no significant difference in CRP between genotypes was reported, except for the Suk et al. study, which showed a significant difference between AA and AT individuals (0.4mg/L, p<0.00001) and between AA and TT individuals (0.5mg/L, p=0.005). In per allele analyses, there was an increase in CRP concentration of 0.14mg/L per additional T-allele (p=0.002).

When the differences between studies in the effect of AA and AT genotype on CRP concentration were examined, there were low levels of heterogeneity (Q=9.164, I^2=67.26%, p=0.03). When the differences between studies in the effect of AA and TT genotype on CRP concentration were examined, there was no evidence of significant heterogeneity (Q=4.067, I^2=26.24%, p=0.25). A funnel plot for the AA vs. AT comparison was then constructed to determine if bias or heterogeneity was present in the studies included for meta-analysis (see Figure 6.8). The Suk et al. study was found to lie just outside the funnel and the PHS cohort from the Miller et al. study was on the border, although overall the funnel appeared symmetrical using a random effects model. The Egger test gave a p value of 0.247, which was not significant for presence of bias.

Figure 6.8. Funnel plot to show presence of bias in the studies used in the meta-analysis to compare CRP concentration between +194AA and +194AT individuals.

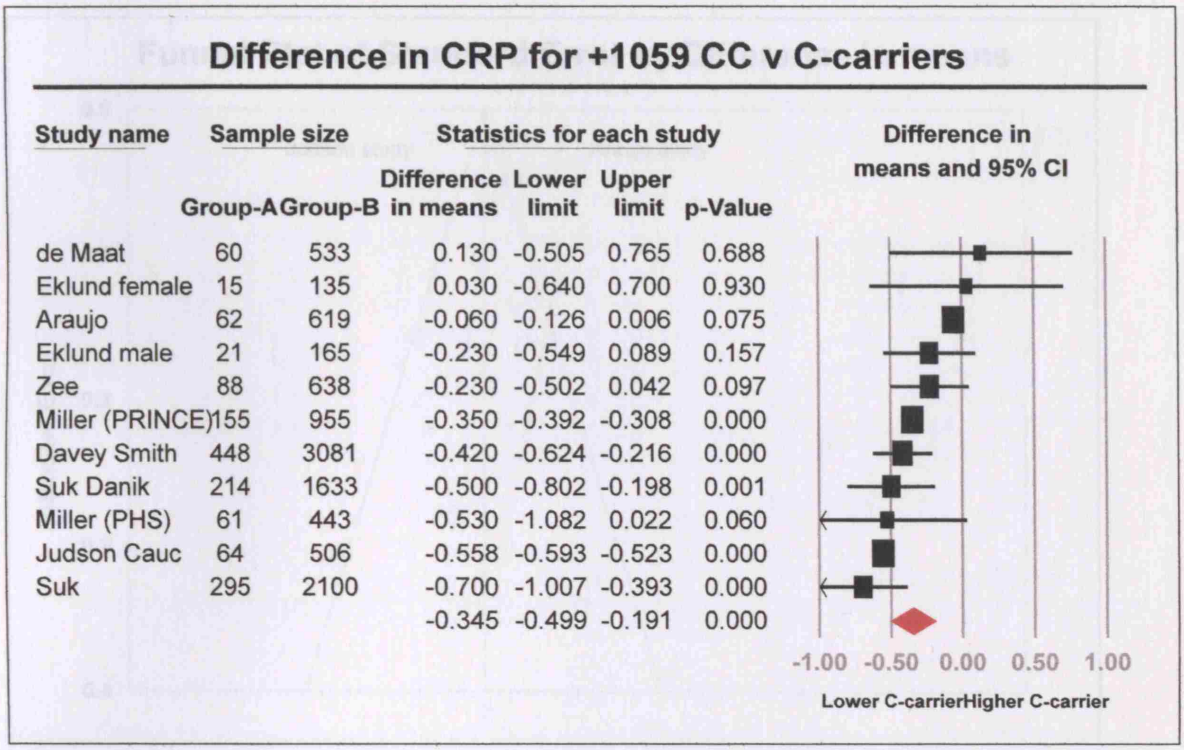


It was not possible to examine the effect of this polymorphism on other intermediate phenotypes as only the study by Suk et al. had available data for other measures, out of which only smoking status and LDL cholesterol showed small associations with genotype (p=0.02 and 0.03 respectively) (Suk *et al.* 2006).

6.4.6 +1059G/C polymorphism and CRP concentration

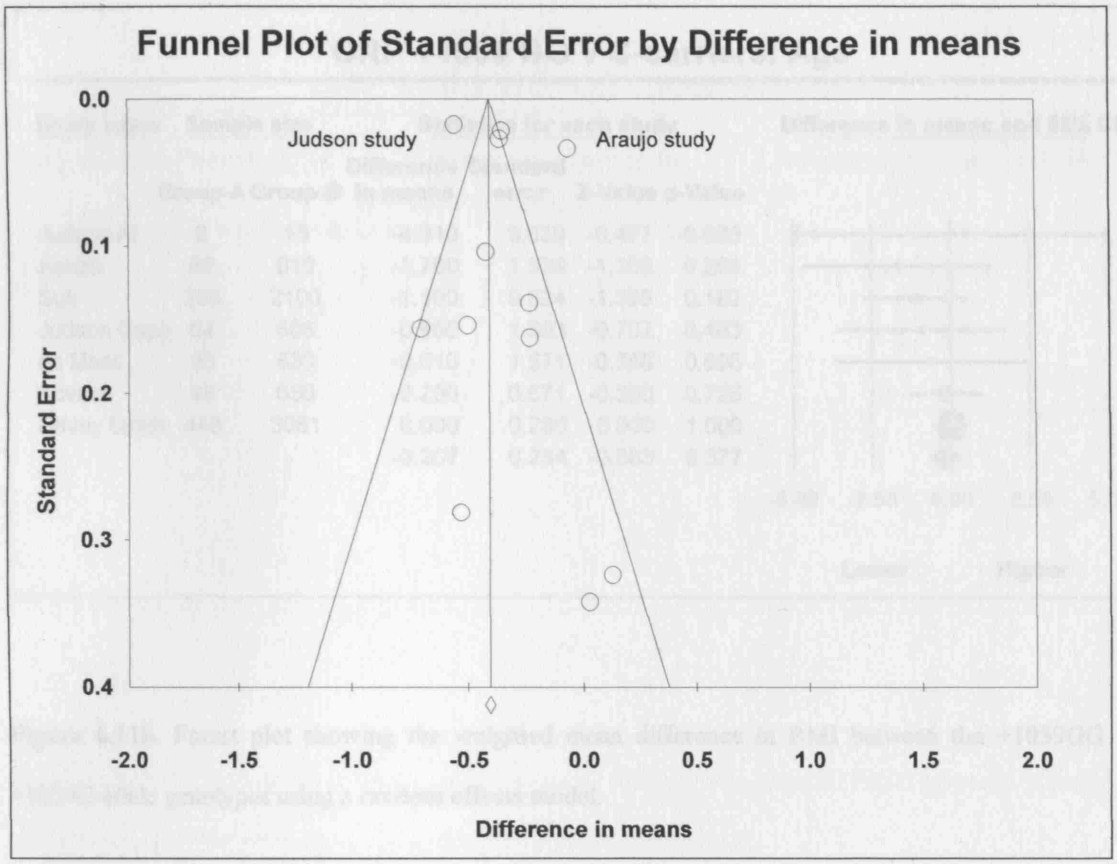
Data from 11 studies totalling 13025 individuals were used for this analysis (Suk Danik *et al.* 2006; Miller *et al.* 2005; Eklund *et al.* 2005; Davey Smith & Ebrahim 2003; Suk *et al.* 2006; de Maat *et al.* 2004; Zee & Ridker 2002; Araujo *et al.* ; Judson *et al.* 2004; Kovacs *et al.* 2004). The majority of studies had a larger percentage of males compared to females, with the exception of the Davey Smith et al. study, which comprised all female subjects, the de Maat et al. study, which had only 32.5% male subjects, and the subjects from the Eklund et al. study, where it was possible to separate the data according to gender. Most of the subjects were Caucasian, with the exception of a few studies that had a mixed population and it was not possible to separate the data according to ethnicity (Suk *et al.* 2006; Araujo *et al.*). The frequencies of the alleles were assessed by χ^2 analysis and were in Hardy-Weinberg equilibrium. Since the CC genotype had a low frequency of approximately 0.04, the GC and CC genotypes were combined. The CRP concentrations were compared by meta-analysis for the genotypes GG vs. C-allele carriers and the summary data were obtained and forest plots were constructed (see Figure 6.9).

Figure 6.9. Forest plot showing the weighted mean difference in CRP concentration between the +1059GG vs. +1059C-allele genotypes using a random effects model.



These results suggest that individuals that carry the C-allele have a CRP concentration 0.348mg/L lower than individuals with the +1059GG genotype ($p<0.0001$) using a random effects model. When the differences between studies in the effect of genotype on CRP concentration were examined, there were significant levels of heterogeneity (Q=200.57, I^2=94.52%, $p<0.00001$). These high levels of heterogeneity could be partly attributable to the studies by Araujo et al. and Judson et al. When these studies were omitted, the heterogeneity values were not statistically significant (Q=11.52, I^2=21.88%, $p=0.242$). A funnel plot was then constructed to determine if bias or heterogeneity was present in the studies included for meta-analysis (see Figure 6.10). The studies by Araujo et al. and Judson et al. were found to lie outside the funnel, although overall the funnel appeared symmetrical when using a random effects model. The Egger test gave a p value of 0.492, which was not significant for presence of bias.

Figure 6.10. Funnel plot to show presence of bias in the studies used in the meta-analysis to compare CRP concentration between +1059GG and +1059C-allele carrier individuals.



A number of other intermediate phenotypes were also available for comparison by genotype, including age, BMI, blood pressure, cholesterol and glucose levels. All comparisons were carried out between +1059GG individuals and +1059C-allele carriers (see Figure 6.11). No significant differences were seen by genotype for any of the variables studied, with the exception of HDL cholesterol, where C-allele carriers had 0.037mmol/L lower levels compared to GG homozygous individuals (p=0.005). When the individual studies were examined, only the Caucasian subjects from the Judson et al. study showed significant differences in HDL cholesterol levels according to genotype (p=0.011). There was no evidence of heterogeneity between studies for any of the comparisons (p=0.144-0.979).

Figure 6.11a. Forest plot showing the weighted mean difference in age between the +1059GG vs. +1059C-allele genotypes using a random effects model.

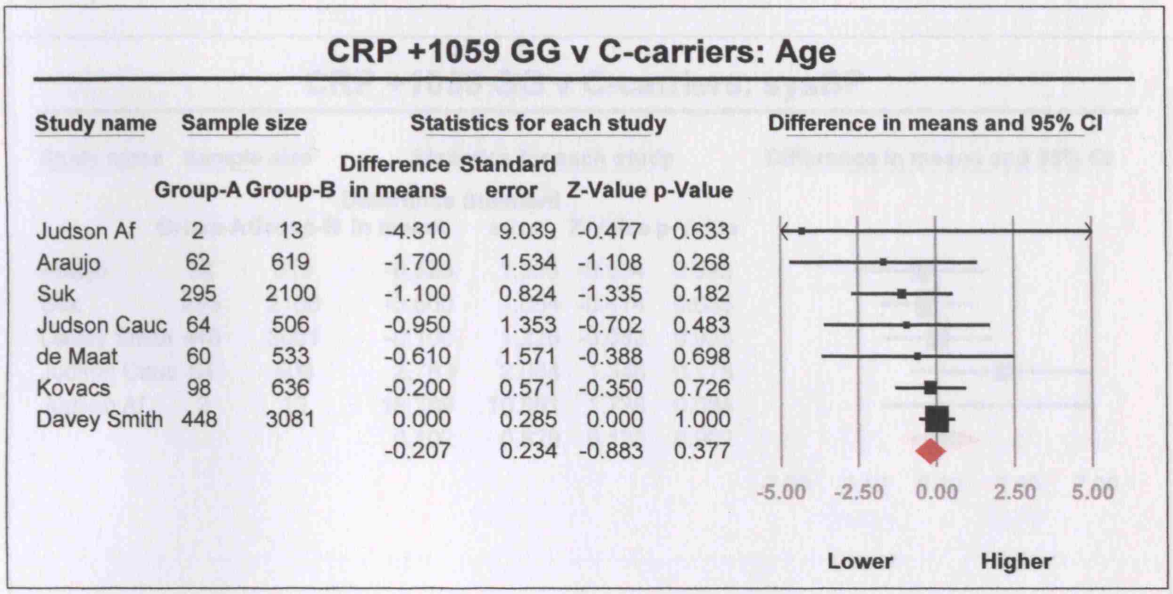


Figure 6.11b. Forest plot showing the weighted mean difference in BMI between the +1059GG vs. +1059C-allele genotypes using a random effects model.

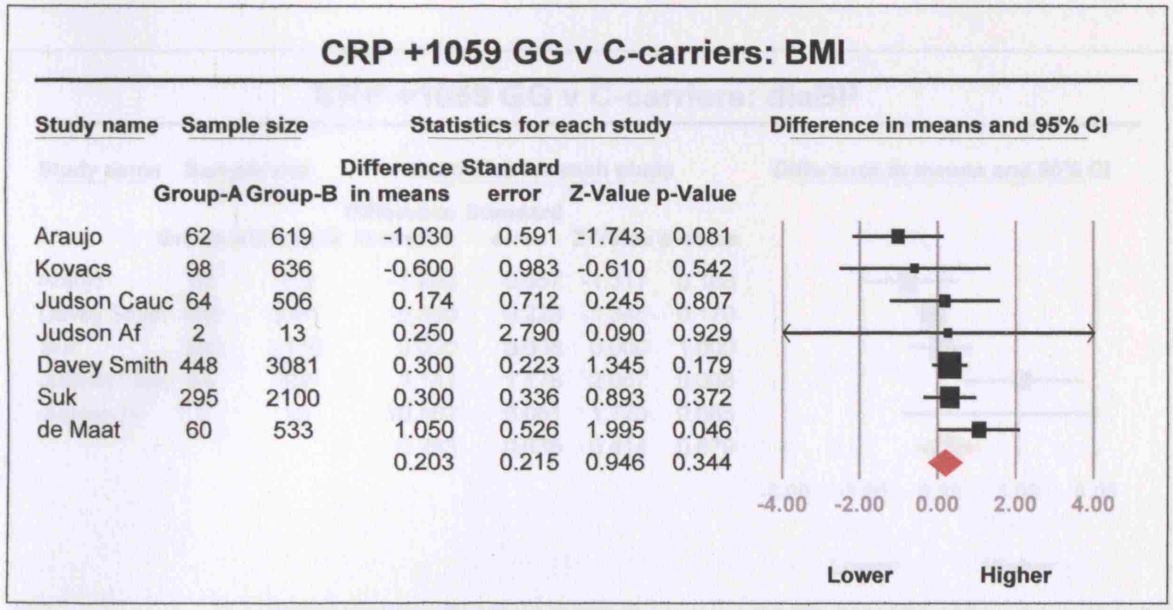


Figure 6.11c. Forest plot showing the weighted mean difference in systolic blood pressure between the +1059GG vs. +1059C-allele genotypes using a random effects model.

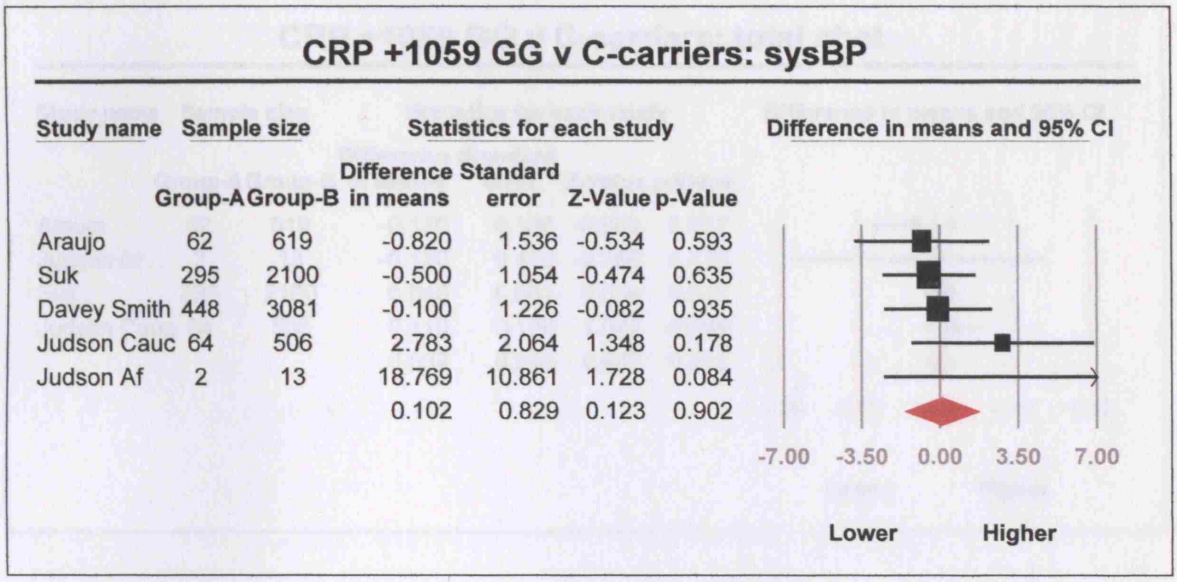


Figure 6.11d. Forest plot showing the weighted mean difference in diastolic blood pressure between the +1059GG vs. +1059C-allele genotypes using a random effects model.

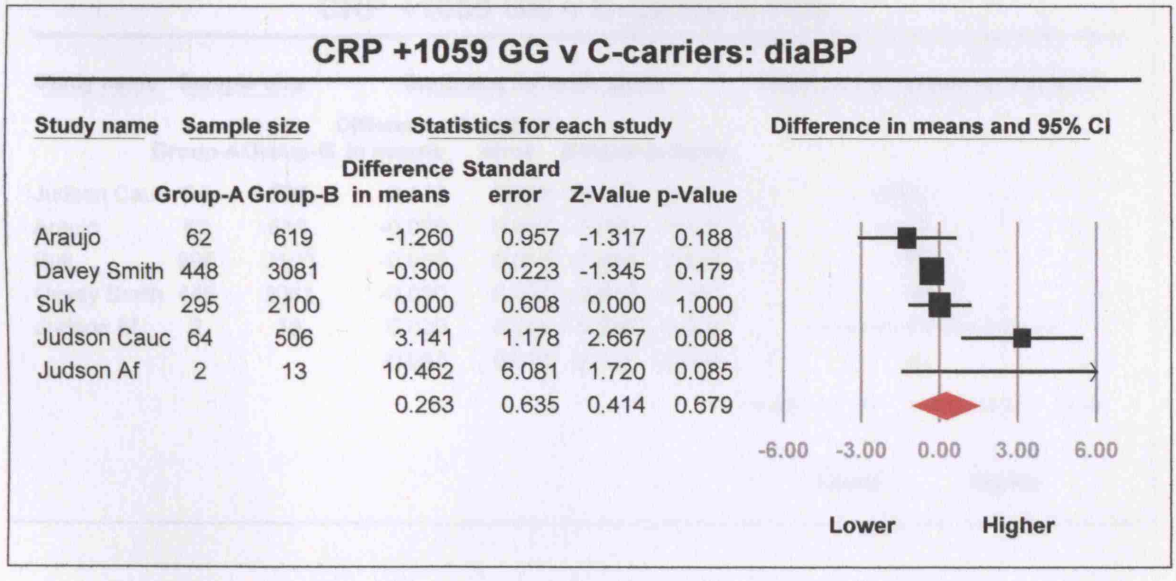


Figure 6.11e. Forest plot showing the weighted mean difference in total cholesterol levels between the +1059GG vs. +1059C-allele genotypes using a random effects model.

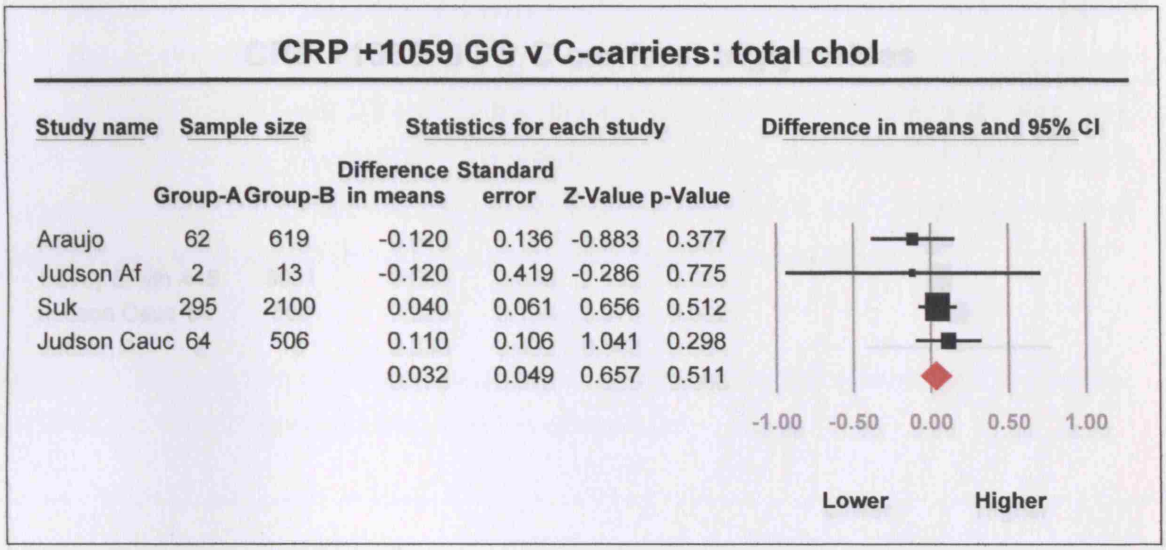


Figure 6.11f. Forest plot showing the weighted mean difference in HDL cholesterol levels between the +1059GG vs. +1059C-allele genotypes using a random effects model.

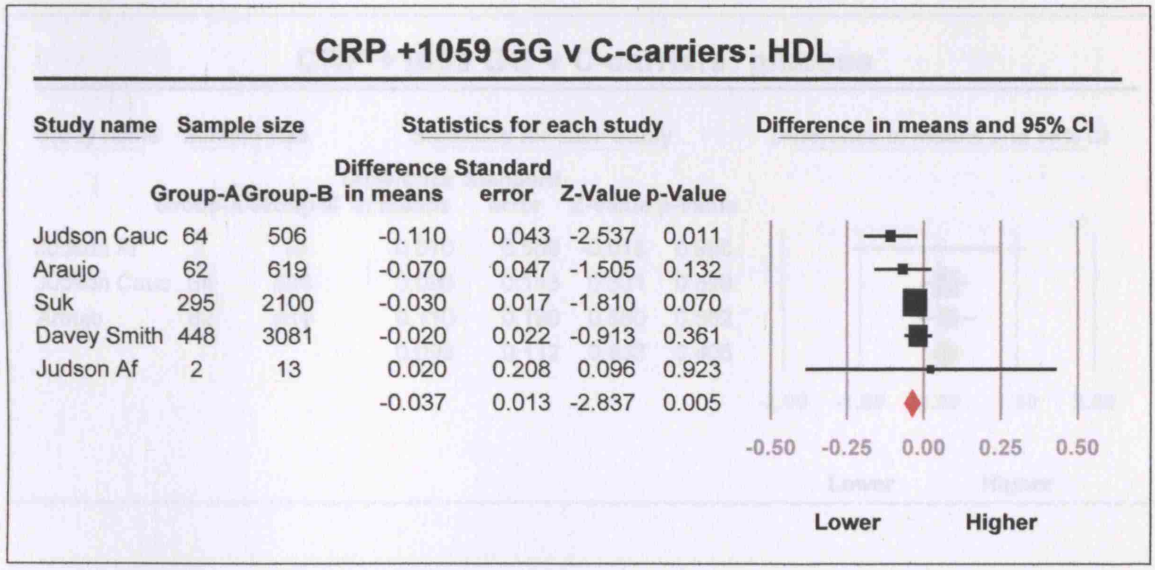


Figure 6.11g. Forest plot showing the weighted mean difference in triglyceride levels between the +1059GG vs. +1059C-allele genotypes using a random effects model.

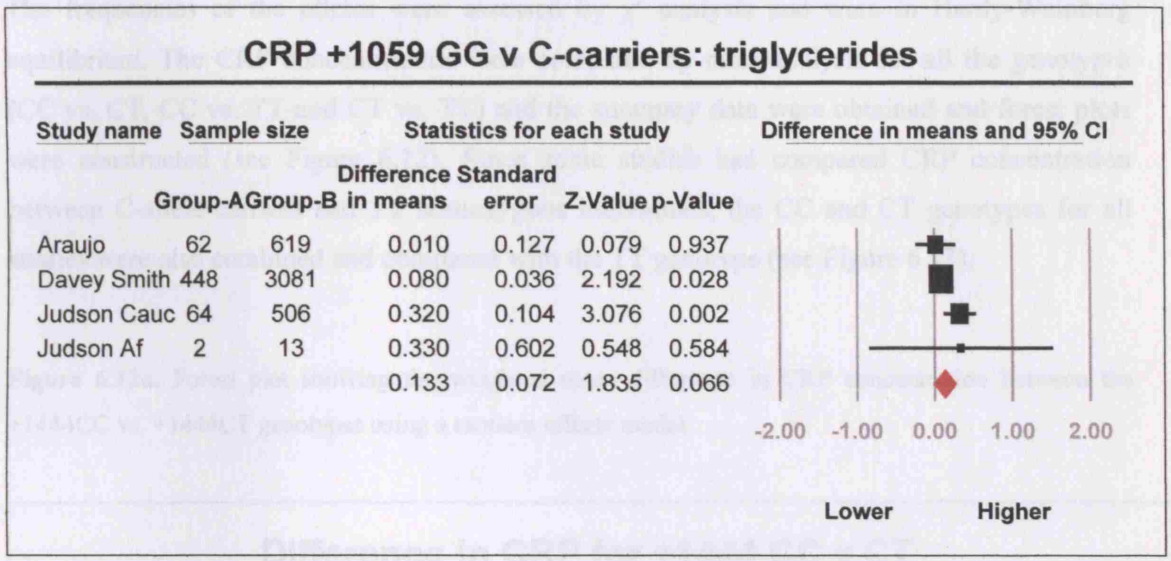
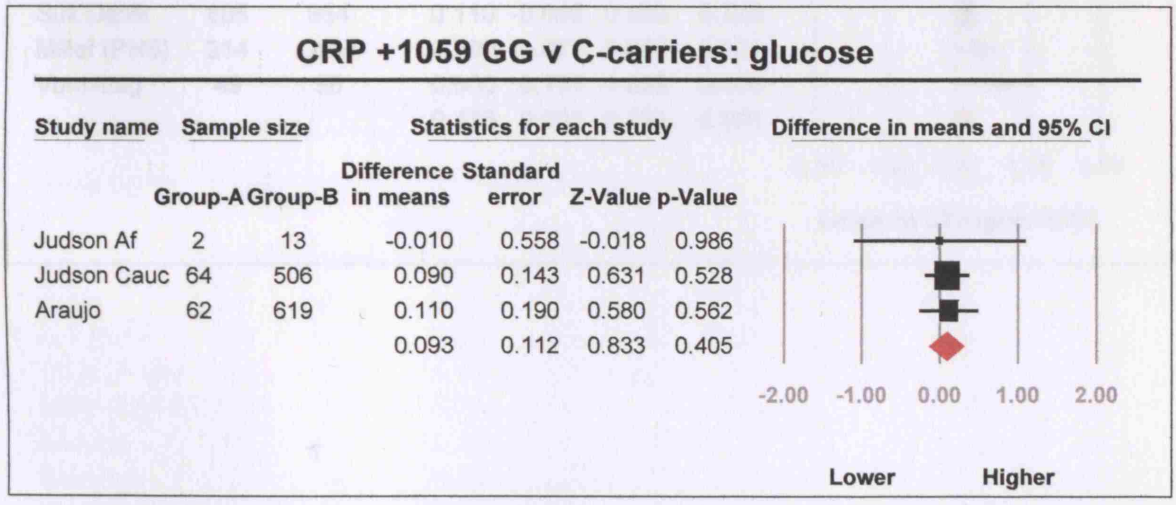


Figure 6.11h. Forest plot showing the weighted mean difference in glucose levels between the +1059GG vs. +1059C-allele genotypes using a random effects model.



6.4.7 +1444C/T polymorphism and CRP concentration

Data from 7 studies involving 6230 individuals were used for this analysis (Willot *et al.* 2006c; Suk Danik *et al.* 2006; Vormittag *et al.* ; Miller *et al.* 2005; Kovacs *et al.* 2004; Brull *et al.* 2003). The coronary artery bypass graft (CABG) surgery patients from the Brull et al. study

were omitted, as the analysis was restricted to apparently healthy individuals. The majority of subjects in all studies were male (>61.4%) except for the Vormittag et al. study where there was approximately the same number of males as females, and all studies had Caucasian subjects. The frequencies of the alleles were assessed by χ^2 analysis and were in Hardy-Weinberg equilibrium. The CRP concentrations were compared by meta-analysis for all the genotypes (CC vs. CT, CC vs. TT and CT vs. TT) and the summary data were obtained and forest plots were constructed (see Figure 6.12). Since some studies had compared CRP concentration between C-allele carriers and TT homozygous individuals, the CC and CT genotypes for all studies were also combined and compared with the TT genotype (see Figure 6.13).

Figure 6.12a. Forest plot showing the weighted mean difference in CRP concentration between the +1444CC vs. +1444CT genotypes using a random effects model.

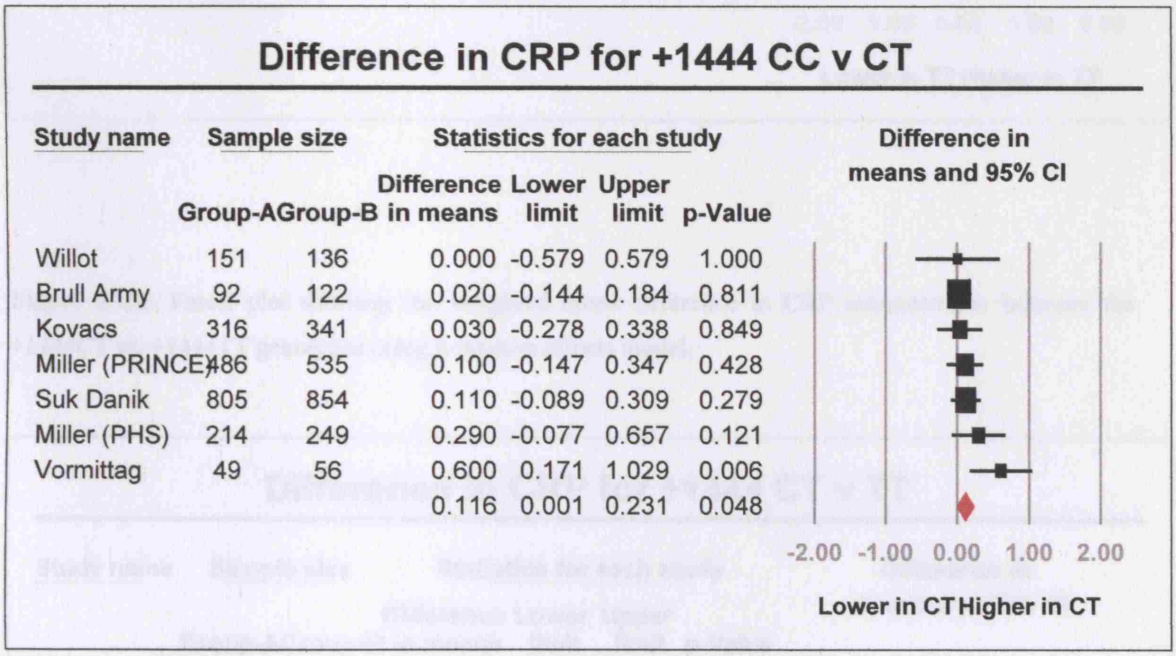


Figure 6.12b. Forest plot showing the weighted mean difference in CRP concentration between the +1444CC vs. +1444TT genotypes using a random effects model.

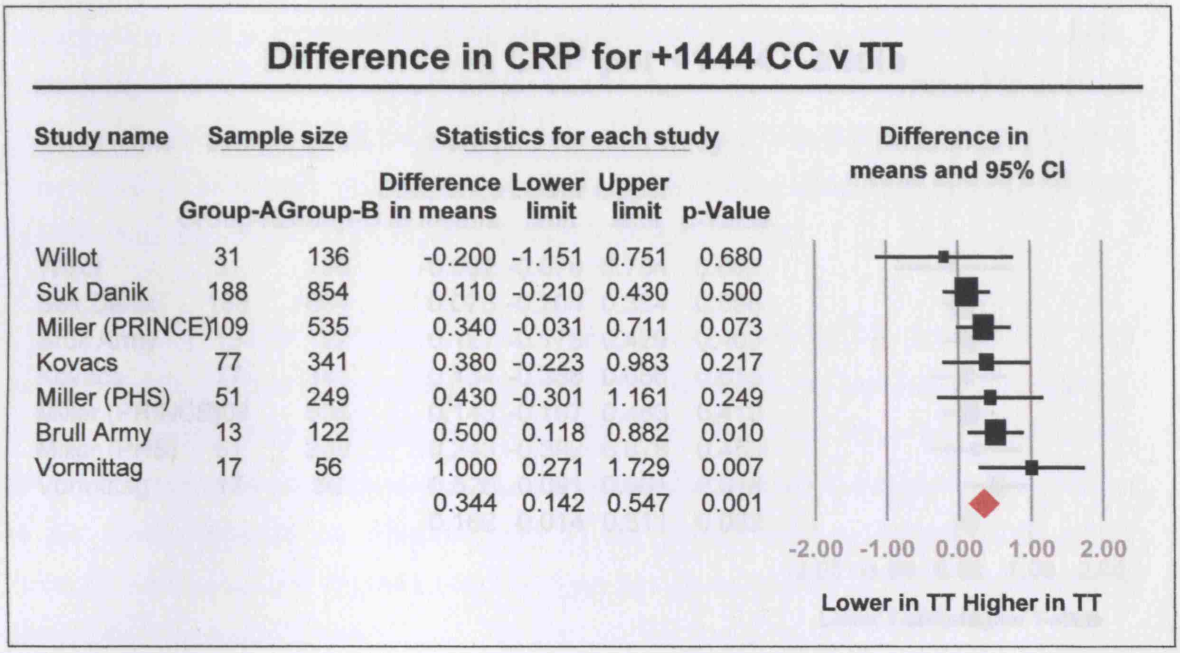


Figure 6.12c. Forest plot showing the weighted mean difference in CRP concentration between the +1444CT vs. +1444TT genotypes using a random effects model.

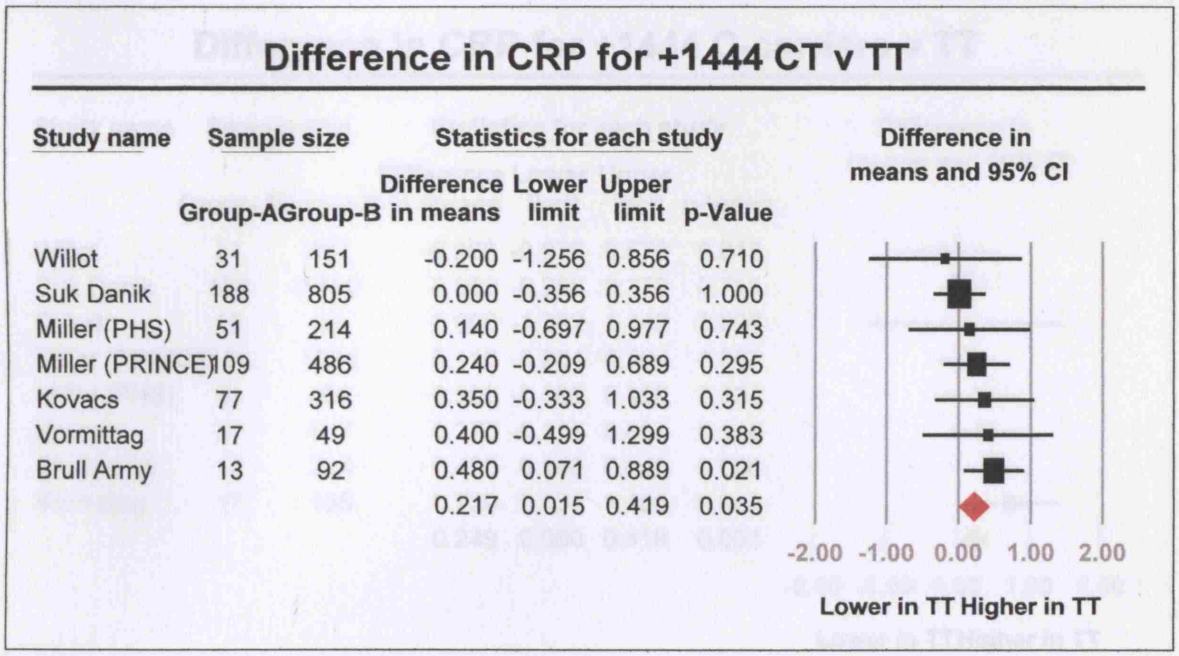


Figure 6.12d. Forest plot showing the weighted mean difference in CRP concentration in per allele analyses.

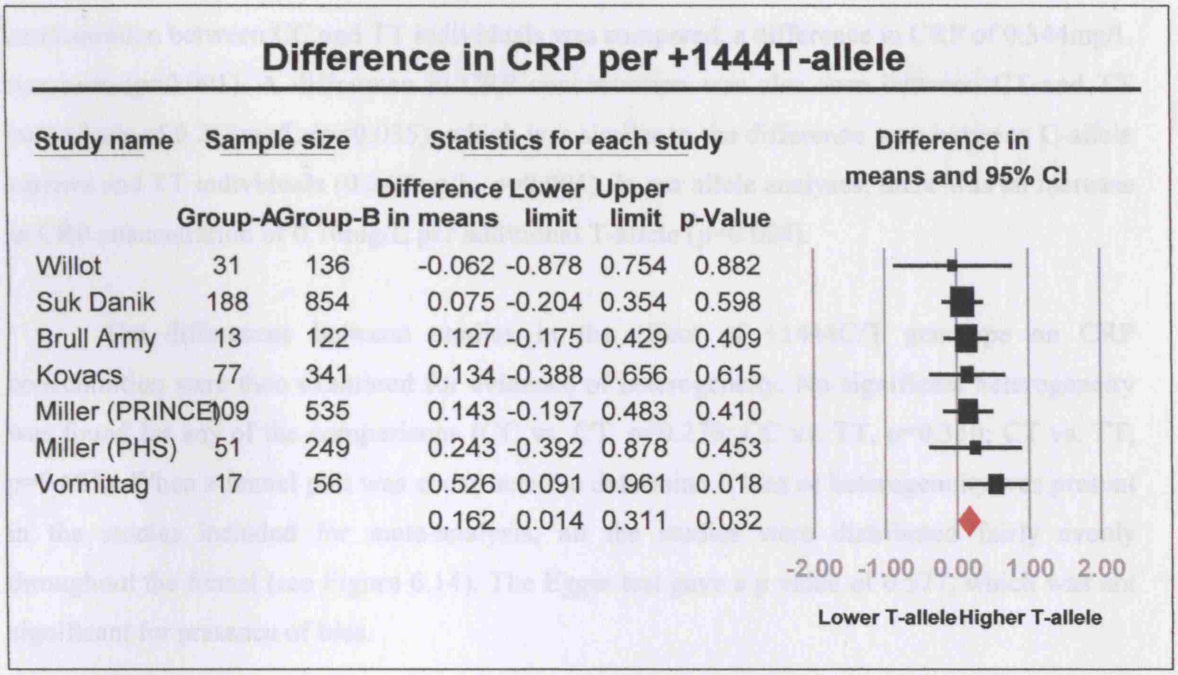
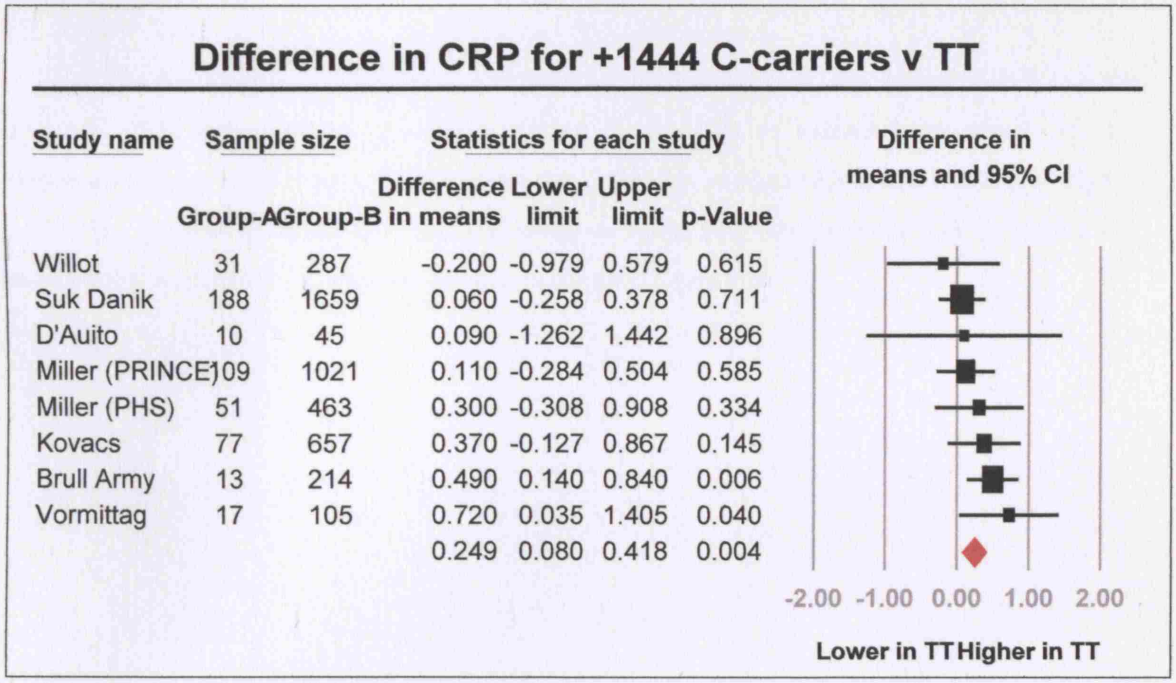


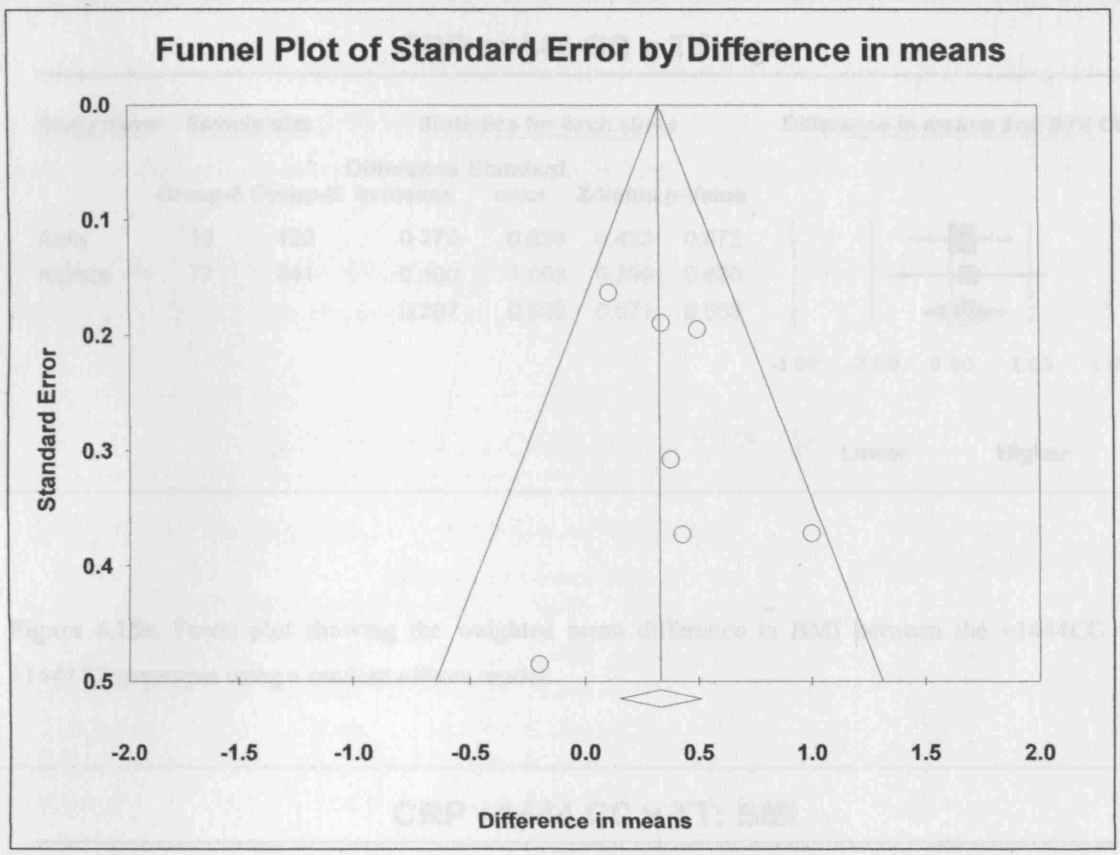
Figure 6.13. Forest plot showing the weighted mean difference in CRP concentration between the +1444 C-allele carrier vs. +1444TT genotypes using a random effects model.



These results suggest that there is a very small significant difference in CRP concentration of 0.116mg/L between individuals with a +1444CC genotype and those with a +1444CT genotype (p=0.048) using a random effects model. When the mean CRP concentration between CC and TT individuals was compared, a difference in CRP of 0.344mg/L was seen (p=0.001). A difference in CRP concentration was also seen between CT and TT individuals of 0.217mg/L (p=0.035), which was similar to the difference seen between C-allele carriers and TT individuals (0.249mg/L, p=0.004). In per allele analyses, there was an increase in CRP concentration of 0.16mg/L per additional T-allele (p=0.004).

The differences between studies in the effect of +1444C/T genotype on CRP concentration were then examined for evidence of heterogeneity. No significant heterogeneity was found for any of the comparisons (CC vs. CT, p=0.279; CC vs. TT, p=0.310; CT vs. TT, p=0.681). When a funnel plot was constructed to determine if bias or heterogeneity was present in the studies included for meta-analysis, all the studies were distributed fairly evenly throughout the funnel (see Figure 6.14). The Egger test gave a p value of 0.577, which was not significant for presence of bias.

Figure 6.14. Funnel plot to show presence of bias in the studies used in the meta-analysis to compare CRP concentration between +1444CC and +1444TT individuals.



The Army study by Brull et al. and the study by Kovacs et al. also had available data for age and BMI according to genotype, making it possible to examine the effect of this polymorphism on other intermediate phenotypes, although no association was found (see Figure 6.15). The Brull et al. study also had measures on blood pressure (systolic and diastolic) and cholesterol, none of which showed any association with genotype.

Figure 6.15a. Forest plot showing the weighted mean difference in age between the +1444CC vs. +1444TT genotypes using a random effects model.

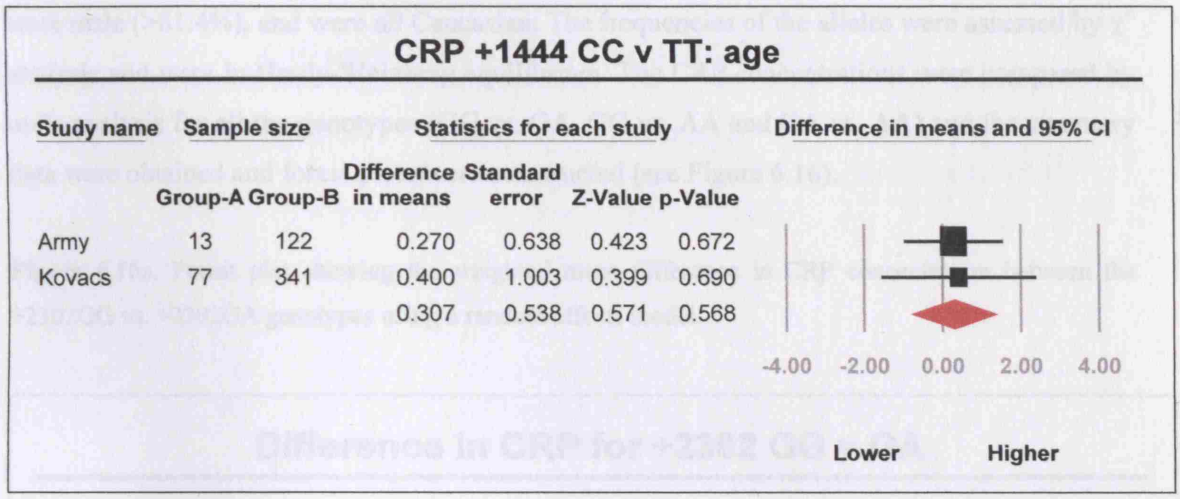
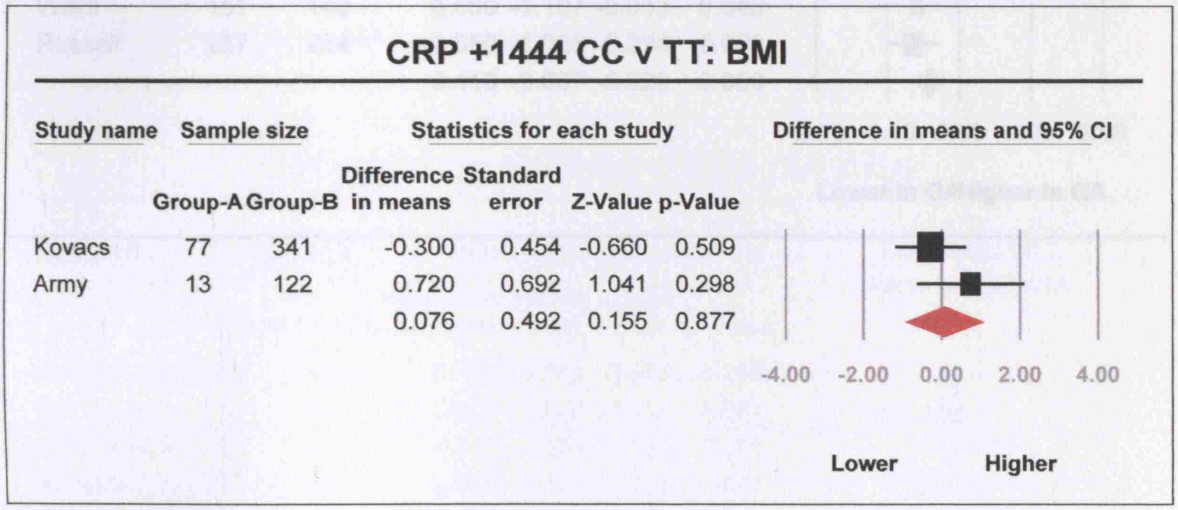


Figure 6.15b. Forest plot showing the weighted mean difference in BMI between the +1444CC vs. +1444TT genotypes using a random effects model.



6.4.8 +2302G/A polymorphism and CRP concentration

Data from 5 studies totalling 4316 individuals were used for this analysis (Willot *et al.* 2006b; Suk Danik *et al.* 2006; Miller *et al.* 2005; Russell *et al.* 2003). The majority of subjects were male (>61.4%), and were all Caucasian. The frequencies of the alleles were assessed by χ^2 analysis and were in Hardy-Weinberg equilibrium. The CRP concentrations were compared by meta-analysis for all the genotypes (GG vs. GA, GG vs. AA and GA vs. AA) and the summary data were obtained and forest plots were constructed (see Figure 6.16).

Figure 6.16a. Forest plot showing the weighted mean difference in CRP concentration between the +2302GG vs. +2302GA genotypes using a random effects model.

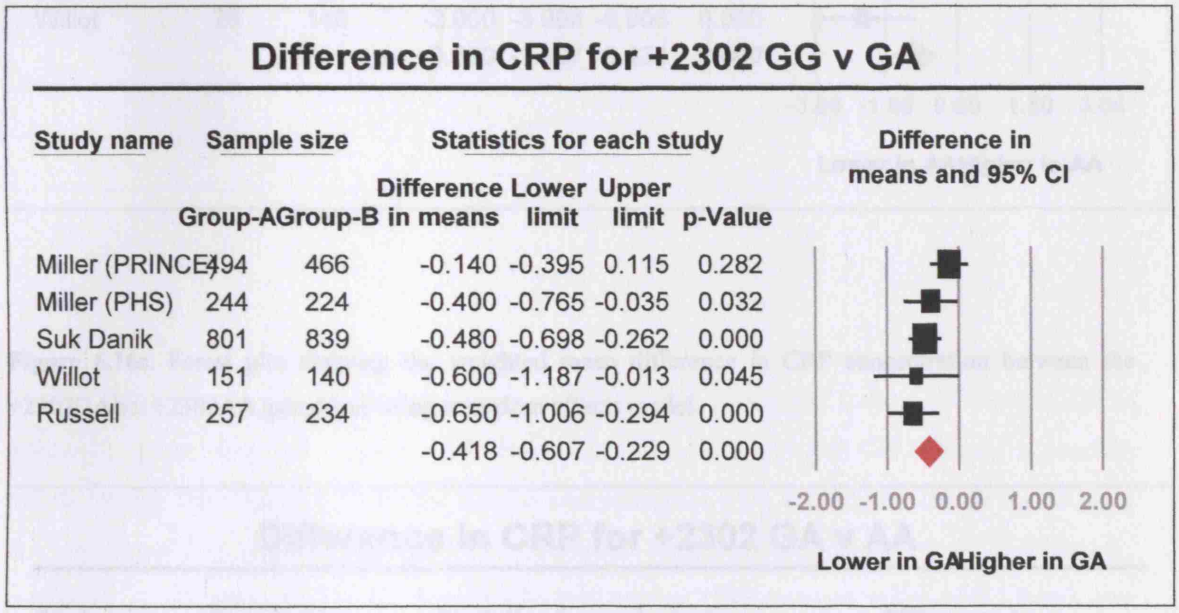


Figure 6.16b. Forest plot showing the weighted mean difference in CRP concentration between the +2302GG vs. +2302AA genotypes using a random effects model.

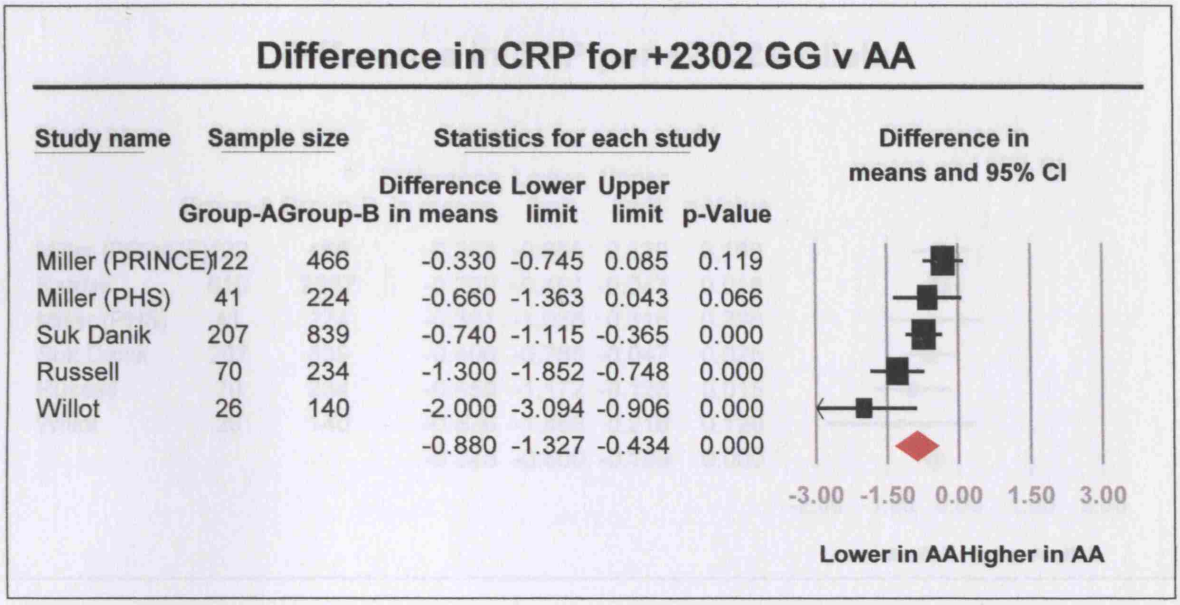


Figure 6.16c. Forest plot showing the weighted mean difference in CRP concentration between the +2302GA vs. +2302AA genotypes using a random effects model.

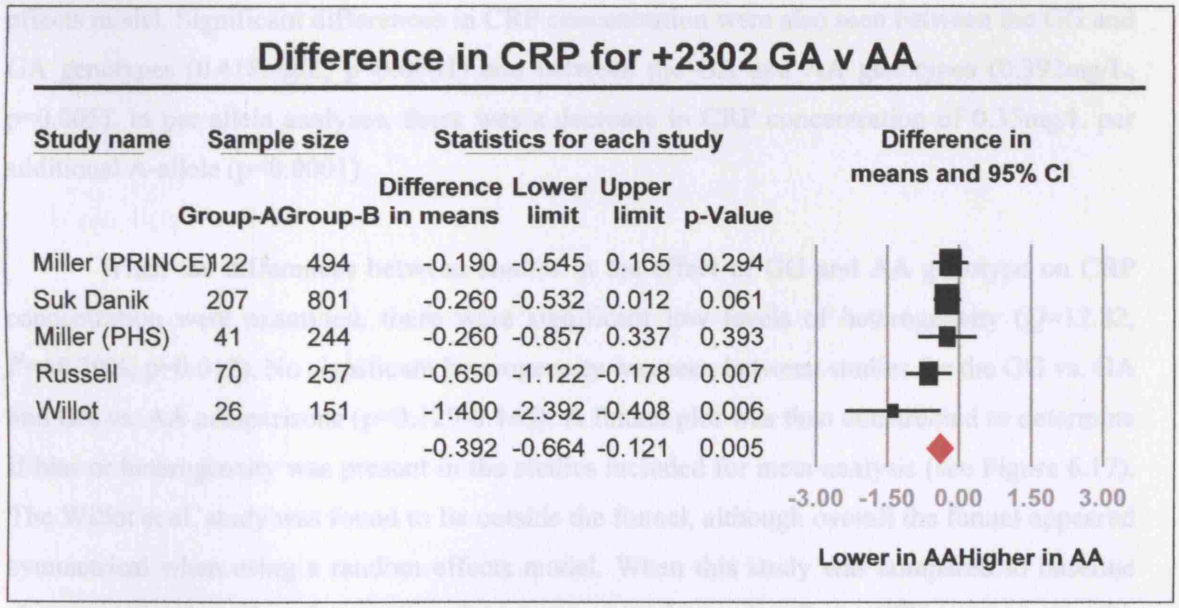
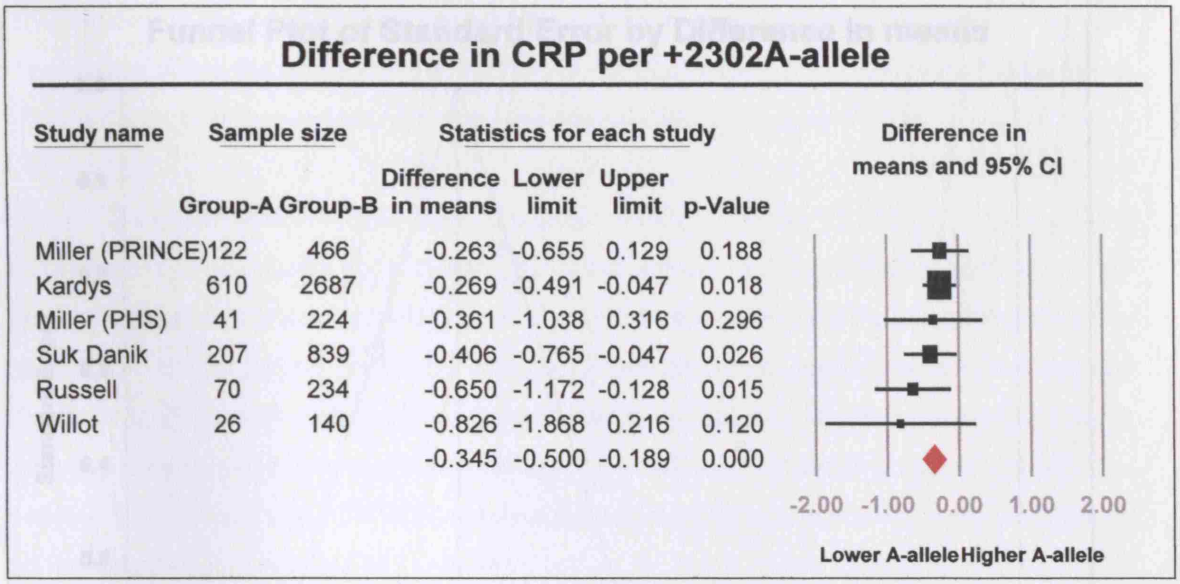


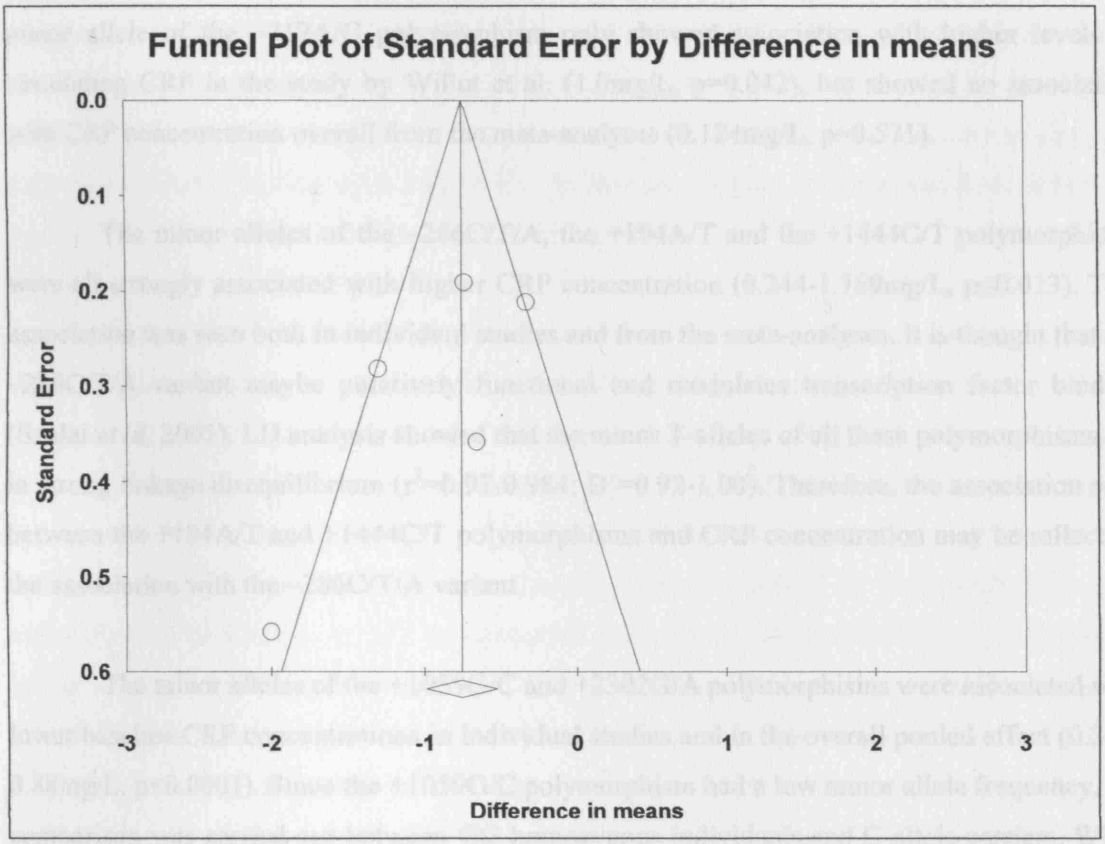
Figure 6.16c. Forest plot showing the weighted mean difference in CRP concentration in per allele analyses.



These results suggest that individuals with the AA genotype have a CRP concentration 0.88mg/L lower than individuals with the +2302GG genotype ($p<0.0001$) using a random effects model. Significant differences in CRP concentration were also seen between the GG and GA genotypes (0.418mg/L, $p<0.0001$) and between the GA and AA genotypes (0.392mg/L, $p=0.005$). In per allele analyses, there was a decrease in CRP concentration of 0.35mg/L per additional A-allele ($p<0.0001$).

When the differences between studies in the effect of GG and AA genotype on CRP concentration were examined, there were significant low levels of heterogeneity ($Q=12.82$, $I^2=68.79\%$, $p=0.012$). No significant heterogeneity was seen between studies for the GG vs. GA and GA vs. AA comparisons ($p=0.129$-0.145). A funnel plot was then constructed to determine if bias or heterogeneity was present in the studies included for meta-analysis (see Figure 6.17). The Willot et al. study was found to lie outside the funnel, although overall the funnel appeared symmetrical when using a random effects model. When this study was compared to baseline characteristics from the other studies, there appeared to be no significant differences, except for sample size, which was much smaller in the Willot et al. study, and may account for the heterogeneity. The Egger test gave a p value of 0.215, which was not significant for presence of bias.

Figure 6.17. Funnel plot to show presence of bias in the studies used in the meta-analysis to compare CRP concentration between +2302GG and +2302AA individuals.



The effect of this polymorphism on other intermediate phenotypes could not be examined, as other measures apart from CRP were not available.

6.5 Discussion

Resequencing and genotype analysis of the CRP gene has led to the identification of several common polymorphisms. Many of these polymorphisms have been the focus of studies that have aimed to determine whether these genetic variants are associated with circulating CRP concentrations. Data from these studies were pooled in meta-analyses for six common polymorphisms to determine precise estimates for the effect on CRP concentration in predominantly male, Caucasian populations. The polymorphisms studied were the –717A/G SNP (rs2794521), the -286C/T/A SNP (rs3091244), the +194A/T SNP (rs1417938), the +1059G/C SNP (rs1800947), the +1444C/T SNP (rs1130864), and the +2302G/A SNP (rs1205). The LD between these variants was also examined from studies where more than one SNP had been studied. In addition, where possible, the effect of these polymorphisms on other intermediate phenotypes was also examined.

The pooled estimates of the association between the minor alleles of five of these polymorphisms and higher or lower CRP concentrations were consistent with the results seen in individual studies, although the size of the effect varied substantially in individual studies. The minor allele of the –717A/G polymorphism only showed association with higher levels of circulating CRP in the study by Willot et al. (1.0mg/L, p=0.042), but showed no association with CRP concentration overall from the meta-analyses (0.124mg/L, p=0.571).

The minor alleles of the –286C/T/A, the +194A/T and the +1444C/T polymorphisms were all strongly associated with higher CRP concentration (0.244-1.769mg/L, $p \leq 0.013$). This association was seen both in individual studies and from the meta-analyses. It is thought that the –286C/T/A variant maybe putatively functional and modulates transcription factor binding (Szalai *et al.* 2005). LD analysis showed that the minor T-alleles of all these polymorphisms are in strong linkage disequilibrium (r^2=0.97-0.984; D'=0.92-1.00). Therefore, the association seen between the +194A/T and +1444C/T polymorphisms and CRP concentration may be reflecting the association with the –286C/T/A variant.

The minor alleles of the +1059G/C and +2302G/A polymorphisms were associated with lower baseline CRP concentrations in individual studies and in the overall pooled effect (0.348-0.88mg/L, p<0.0001). Since the +1059G/C polymorphism had a low minor allele frequency, the comparison was carried out between GG homozygous individuals and C-allele carriers. When the LD between these two SNPs was examined, only weak LD as measured by r^2 was seen (r^2=0.086-0.140). This may be because of different minor allele frequencies (MAF), and suggests that each polymorphism may contribute to the effect on CRP concentration, possibly in an additive fashion. However, when the LD was examined using the D' measure, which is less sensitive to differences in MAF, higher LD was seen (D'=0.758-0.979), suggesting that these two polymorphisms lie within the same haplotype block. One explanation for the association seen with CRP concentration may be that the +2302G/A variant, which lies in the 3' UTR, may influence mRNA stability and thus influence CRP concentration. Therefore, the association seen with the +1059G/C variant may be due to the LD with the +2302G/A polymorphism.

Data on other intermediate phenotypes including age, BMI, blood pressure, cholesterol and glucose were available for comparison by genotype from some of the studies that examined genotype-CRP association. It was possible to evaluate the relationship between genotype and these variables for the +1059G/C polymorphism and between genotype and age and BMI for the +1444C/T polymorphism. For the +1059G/C polymorphism, all comparisons were carried out between +1059GG individuals and +1059C-allele carriers and no significant differences were seen by genotype for any of the variables studied, with the exception of HDL cholesterol, where C-allele carriers had 0.037mmol/L lower levels compared to GG homozygous individuals

(p=0.005). When the individual studies were examined, only the Caucasian subjects from the Judson et al. study showed significant differences in HDL cholesterol levels according to genotype (p=0.011). This significant association may simply be due to chance since a number of comparison tests were carried out. Similarly, for the +1444C/T polymorphism, no association was seen between genotype and age or BMI. This suggests that the effect of CRP genotype is not confounded by other variables that are associated with CRP concentration since genotype is allocated randomly at conception according to Mendel's Laws, so other variables should be evenly distributed among the different genotypic groups.

Heterogeneity between studies was examined when each meta-analysis was conducted using the DerSimonian and Laird Q test, and the I^2 test (Higgins *et al.* 2003). In some analyses, a significant level of heterogeneity was seen, particularly when an individual study had much larger differences in CRP concentration by genotype compared to the overall weighted mean difference. One explanation for this may be due to ethnicity differences in the cohorts studied. An example of this can be seen when investigating the –286C/T/A polymorphism, where unusually high differences in CRP concentration can be seen between the CC and TT genotype in the African cohort of the Szalai et al. study (8.624mg/L; p<0.0001). The other studies included in this meta-analysis had predominantly Caucasian subjects, whereas the Szalai et al. study was stratified according to ethnicity. Another possibility may be differences in CRP measurement, leading to a greater variation in the genotype-CRP association and thus resulting in statistical heterogeneity. Where significant levels of heterogeneity were observed, the statistical tests were repeated with studies omitted to determine the source of heterogeneity.

One limitation in conducting a systematic review of CRP genotype and CRP concentration associations is combining data of different SNPs that have been typed. Although the LD structure of the CRP gene is known and therefore it is possible to gain information on untyped SNPs from genotyped SNPs, there is no formal approach to combine data on different typed SNPs that are in LD with each other. The work conducted in this chapter provided the motivation for collaboration with John Whittaker and Claudio Verzilli (London School of Hygiene and Tropical Medicine) to develop such methods.

6.6 Conclusions

Meta-analyses of published data were carried out to obtain precise estimates of the effect of common variation in the CRP gene on CRP concentration. Searches on three electronic databases identified 27 publications on CRP polymorphisms and CRP concentration. From these publications, 12 were included for data extraction and genotype-CRP analysis (14 studies). Data from 8 of these publications were also used to examine the effect of genotype on

other intermediate phenotypes. A total of 6 polymorphisms were assessed, from which 5 showed association with CRP concentration. The minor alleles of the –286C/T/A, +194A/T and +1444C/T polymorphisms were all associated with higher CRP concentration. The minor alleles of the +1059G/C and +2302G/A polymorphisms were associated with lower CRP concentration. The effect of genotype on CRP concentration did not appear to be confounded by other variables also associated with CRP concentration. These estimated effects were in concordance with individual SNP study results with known effects on CRP.

7. Effect of CRP genotype and haplotype on basal CRP levels

7.1 Aim

To determine if polymorphisms within the CRP gene, and haplotypes generated using CRP tagging SNPs are robustly associated with differences in basal CRP concentrations in healthy individuals. To test whether the association is modified by a variety of exposures relevant to coronary risk. To confirm the randomised allocation of alleles lead to the expected balanced distribution of covariables by genotype and haplotype class.

7.2 Background

The association of circulating concentrations of CRP with later cardiovascular events in several prospective observational studies led to the suggestion that CRP measurement might have utility in disease prediction (Heeschen *et al.* 2000; Zebrack *et al.* 2002; Ridker & Haughie 1998; Ridker *et al.* 2002; Danesh *et al.* 2000a). However, when the predictive power of CRP was evaluated in the NPHSII cohort using a variety of appropriate methods (see Chapter 4), CRP did not perform particularly well either in isolation or when evaluated in conjunction with established risk factors. These findings were consistent with a re-analysis of data from prospective studies.

Since the predictive utility of a CRP measure may be limited, the major focus of interest in the association of CRP with cardiovascular disease may be its causal relevance. If CRP plays a pathogenic role in cardiovascular disease risk, then it could be an important therapeutic target and its reduction might help CVD prevention. However, it is only possible to make limited inference on causation from observational data because of the problems of confounding and reverse causality. Studying CRP indirectly via polymorphisms that influence its concentration may provide a better insight on causation because genotype is allocated at random at conception, so established risk factors for cardiovascular disease and other covariables should be distributed evenly across the genotypic classes. Therefore any association of genotype with disease should be free from confounding and reverse causality, allowing inferences on a potential causal role for CRP to be made.

Data in the previous chapter (Chapter 6) provided an overview of published information on the association of several SNPs in the CRP gene with CRP concentrations. Because meta-analyses of published genetic association studies can be affected by publication bias, one aim of the work in this chapter was to confirm the association by undertaking new genotyping in larger

data sets. For this work, a more systematic approach was used that involved the genotyping of tagging SNPs. Based on the work in Chapter 5, three common SNPs out of twelve in the CRP gene were identified in European populations that together capture almost all the haplotype diversity in this population. The SNPs chosen were the +1444C/T, +2302G/A and +4899T/G SNPs. CRP haplotypes were then inferred to allow tests for association of haplotype with CRP concentration. A final aim of this chapter, facilitated by the availability of individual participant data and a wide range of measures of covariables, was to evaluate whether the genotype/haplotype-CRP association was modified among subjects of differing age and prevailing levels of established risk factors and other covariates. Work was also undertaken to confirm the random allocation of alleles at conception reduces confounding by balancing the distribution of established risk factors and other variables among genotypic and haplotypic classes. This information is important to evaluate the validity of Mendelian randomisation analyses.

Previous studies have described associations with CRP of two of the SNPS comprising this trio, the +1444C/T and +2302G/A polymorphisms (see Chapter 6) (Brull *et al.* 2003; Russell *et al.* 2003). However, no previous studies have examined the effect of the +4899T/G variant on CRP levels. It would be expected that if this strategy is efficient, information on any untyped SNPs including a SNP or SNPs with potential functional relevance would be captured indirectly by the tagging SNPs. This is an important issue for whole genome association studies where causative variants are unknown and the use of tagging SNPs has been proposed as an efficient means of capturing genetic variation to infer the location and haplotypes harbouring causal variants by typing only a subset of all SNPs across the human genome. Because a possible functional SNP in the CRP gene has already been proposed (–286C/T/A, rs3091244), typing this SNP as well as the tagging SNPs provided the opportunity to test the validity of a tagging SNP approach using a continuous outcome (CRP concentration), which helps to enhance power over genotype-disease association studies. The studies were undertaken in two population-based cohorts of healthy individuals to reduce the potential modifying effect of disease itself on the genotype/haplotype-CRP association.

7.3 Methods

7.3.1 Study subjects

Detailed descriptions of the two studies are given in Chapter 3. Brief details are reiterated here.

NPHSII cohort

The NPHSII cohort is a large prospective study of 3,012 healthy Caucasian men aged 50 to 61 years, originally recruited in 1986. Nine general practices participated in the study, and all patients were free of a history of unstable angina, MI or evidence of silent infarction, coronary surgery, aspirin or anticoagulant therapy, cardiovascular disease, or malignant disease at the time of recruitment. Baseline characteristics and demographic information were ascertained by means of a clinical assessment and a questionnaire completed at the beginning of the study. The endpoints were fatal CHD events and non-fatal MI, coronary artery surgery and silent MI on follow-up ECG. At present, 227 CHD events have been recorded. Genotyping was carried out on 2,676 men; the rest were non-Europeans or had incomplete measurements and were therefore ineligible.

ELY cohort

The Ely Study is a prospective population-based cohort study of the aetiology and pathogenesis of type 2 diabetes and related metabolic disorders in 1122 volunteers originally recruited in 1990 from people in Ely, Cambridgeshire. At baseline, socioeconomic status (SES) was collected by self-report. Complete data on biochemical and anthropometric variables were available in 839 subjects. Participants attended the laboratory after an overnight fast for clinical assessment. Blood samples were taken for plasma glucose, total serum cholesterol, HDL, and triglyceride measurement Genotyping was carried out in 598 subjects, all of which had CRP measures.

7.3.2 CRP plasma measurements

In the NPHSII study, CRP was measured using commercial assays (R&D Systems or Kordia Life Sciences). Inter-assay and intra-assay coefficients of variations were <6.2% and <1.9% respectively, with a detection limit of 0.1mg/L. In the Ely study, CRP was measured by immunometric assay using an Immulite Autoanalyzer (Diagnostics Products) with an analytical sensitivity limit of 0.1mg/l and inter-assay and intra-assay coefficients of variation less than 8%.

7.3.3 Genotyping the CRP polymorphisms

Three polymorphisms were genotyped in both the NPHSII and Ely cohorts, the +1444C/T, the +2302G/A, and the +4899T/G comprising a trio of tagging SNPs used to generate four common haplotypes in Caucasians with a frequency of 5% or more. These SNPs were chosen as tagging SNPs based on work in Chapter 5. Genotyping in the Ely cohort was conducted in Cambridge with the collaboration of Dr Manjinder Sandhu and Professor Nick Wareham. In addition, the potentially functional –286C/T/A polymorphism (see Chapters 5 and 6) was genotyped in the NPHSII study.

The +1444C/T polymorphism was genotyped by RFLP analysis in the NPHSII cohort using the forward primer 5' AGC TCG TTA ACT ATG CTG GGG CA 3' and reverse primer 5' CTT CTC AGC TCT TGC CTT ATG AGT 3' to amplify the region containing the variant, and the restriction endonuclease *SduI*, which only cleaves the polymorphic site in the presence of the common C-allele. Genotyping was repeated in a subset of 200 subjects using a TaqMan assay by design (Applied Biosystems) as a quality control. For the Ely cohort, the +1444C/T polymorphism was genotyped using a standard TaqMan assay by design using the forward primer 5' GGT CTG GGA GCT CGT TAA CTA TG 3' and the reverse primer 5' TCC AAC TTG AAA AAC AAA ACA CCT CAA 3'.

The +2302G/A polymorphism was genotyped in both cohorts using the TaqMan assay by design with the forward primer 5' CAC CAG TAG CCA TCT TGT TTG C 3' and the reverse primer 5' CCA CTT CCA GTT TGG CTT CTG T 3'. The +4899T/G polymorphism was also genotyped in both cohorts using the TaqMan assay by design with the forward primer 5' TTA TCC TAG GAC AAC TGC CCA CTA 3' and the reverse primer 5' GGA GCT GAA GAG AAG GAA TCC A 3'.

The –286C/T/A polymorphism was genotyped using a pyrosequencing method (see Chapter 3) using the forward primer 5' TGA TTT GGG CTG AAG TAG GTG 3', the reverse primer 5' TGG CTA TCT ATC CTG CGA AAA T 3' and the sequencing primer 5' ACC CAG ATG GCC ACT 3'.

7.3.4 Statistical analysis

These analyses were mainly conducted by Ms. Jackie Cooper. I contributed to the preparation of tabular data and helped to conduct some of the analyses.

Genotype frequencies were compared to those expected under the assumption of Hardy-Weinberg equilibrium by χ^2 analysis. Associations of genotype with CRP and risk factors were tested using one-way analysis of variance (ANOVA) or Kruskal-Wallis tests for continuous variables and χ^2 or Fisher's exact test for categorical variables. In addition, "per allele" effects were determined using regression analysis.

Haplotypes were generated from three tagging SNPs, selected based on their location and availability of genotyping assays (see Chapter 5). LD between SNPs was evaluated in the NPHSII data set using D' and r^2 statistics. The +1444C/T and +4899T/G SNPs are known to be in LD with other common variants in the CRP gene. However, the +2302G/A SNP marks a unique haplotype and doesn't appear to be in LD with known common CRP variants. Haplotype analysis was performed using a maximum likelihood model based on the stochastic-EM

algorithm implemented in the THESIAS program (http://www. genecanvass.org). THESIAS allows the simultaneous estimation of haplotype frequencies and of their associated effects on the phenotype of interest. The haplotype combining the most frequent alleles at each locus was used as the reference. A global p value was calculated using differences in log-likelihood assuming an additive model of haplotype effects. Haplotypes with a frequency of less than 5% were excluded.

7.4 Results

7.4.1 CRP genotypes and haplotypes in the NPHSII study

Four SNPs in the CRP gene were typed in the NPHSII cohort. Allele frequencies are shown in Table 7.1. There was no evidence of deviation from Hardy-Weinberg equilibrium for any of the polymorphisms studied. The three haplotype tagging SNPs identified four common haplotypes (see Figure 5.6 in Chapter 5) referred to as CAT, CGT, TGT, and CGG, (+1444C/T / +2302G/A / +4899T/G) with frequencies 0.33, 0.31, 0.30 and 0.05 respectively (see Table 7.2).

Table 7.1. Allele and genotype frequencies for each polymorphism genotyped in the NPHSII cohort.

Polymorphism	Major allele frequency	Minor allele frequency	Chi squared	p-value for H-W deviation
-286C/T/A	C = 0.608	T = 0.323 A = 0.069	5.380	0.068
+1444C/T	C = 0.691	T = 0.309	2.403	0.121
+2302G/A	G = 0.666	A = 0.334	0.034	0.853
+4899T/G	T = 0.945	G = 0.055	0.556	0.456

Table 7.2. Haplotypes generated from the three tagging SNPS (+1444C/T, +2302G/A and +4899T/G) with a frequency of 5% or more.

Haplotype	Frequency
CAT (Haplotype 1)	0.33
CGT (Haplotype 2)	0.31
TGT (Haplotype 3)	0.30
CGG (Haplotype 4)	0.05

+1444C/T is at position 1, +2302G/A is at position 2 and +4899T/G is at position 3.

The –286C/T/A polymorphism was genotyped by pyrosequencing. Information on genotype for this polymorphism was then added to the three tagging SNPs and new haplotypes were generated to determine whether the –286C/T/A polymorphism lies within the four common haplotypes generated from the three tagging SNPs or whether new common haplotypes are generated (frequency $\geq$0.05). As shown in Table 7.3, the four common haplotypes inferred were unchanged and the additional ten minor haplotypes in aggregate accounted for around 6% of all chromosomes.

Table 7.3. Haplotypes generated from the three tagging SNPs and the –286C/T/A SNP and their frequencies (frequencies less than 0.001 have been excluded and frequencies greater than 0.05 are in bold).

Haplotype	Frequency	SE
CAT C	**0.3079**	0.0012
CAT T	0.0138	0.0010
CAT A	0.0077	0.0006
CGT C	**0.2848**	0.0012
CGT T	0.0194	0.0010
CGT A	0.0057	0.0006
TGT C	0.0099	0.0006
TGT T	**0.2886**	0.0005
TGT A	0.0029	0.0003
CGG C	0.0013	0.0002
CGG T	0.0016	0.0002
CGG A	**0.0501**	0.0002
TAT C	0.0052	0.0005
TAT T	0.0014	0.0005

+1444C/T is at position 1, +2302G/A is at position 2, +4899T/G is at position 3 and –286C/T/A is at position 4.

The LD between these polymorphisms was then examined. The pairwise D' and r^2 LD values are shown in Table 7.4. The three haplotype tagging SNPs (+1444C/T, +2302G/A and +4899T/G) exhibited low levels of allelic association with one another as expected from their choice as haplotype tagging SNPs. The –286C/T/A promoter SNP and +1444C/T SNP in the 3' UTR exhibited a high degree of linkage disequilibrium (r^2=0.78), and similar high levels of LD were also seen between the A-allele of the –286C/T/A polymorphism and the G-allele of the +4899T/G SNP (r^2=0.69), also lying in the 3' UTR.

Table 7.4. LD between the +1444C/T, +2302G/A, +4899T/G and –286C/T/A polymorphisms as calculated by the D' and r^2 LD statistics.

	+1444C>T	+2302G>A	+4899T>G	-286C>T	-286C>A
+1444C>T		0.96	0.90	0.91	0.75
+2302G>A	0.21		1.00	0.85	0.60
+4899T>G	0.02	0.03		1.00	0.94
-286C>T	0.77	0.18	0.03		
-286C>A	0.02	0.01	0.69		

D' LD is on the upper right (in blue) and r^2 LD is on the lower left (in red) of the table.

7.4.2 Association of CRP genotypes and haplotypes with CRP concentration in the NPHSII study

All polymorphisms exhibited association with CRP concentration. In per allele analyses of the +1444C/T polymorphism, CRP concentration increased by 0.13mg/L for each additional T-allele. In similar analyses of the +4899T/G polymorphism, CRP increased by 0.2mg/L for each additional G-allele. In per allele analyses of the +2302G/A polymorphism, CRP concentration decreased by 0.14mg/L for each additional A-allele (see Table 7.5).

Table 7.5. Association of CRP genotypes with CRP concentration (mg/L) in the NPHSII study (those with CRP>20mg/L have been excluded).

SNP	Genotype*			P-value	Change per allele (SE)	P-value
	11	12	22			
+1444C/T	2.26 (2.29)	2.64 (2.67)	2.96 (3.02)	0.0001	0.13 (0.04)	<0.0001
+2302G/A	2.73 (2.69)	2.36 (2.42)	2.05 (2.15)	<0.0001	-0.14 (0.03)	<0.0001
+4899T/G	2.43 (2.46)	3.16 (3.08)	1.70 (2.06)	0.0008	0.20 (0.07)	0.003
-286C/T	2.19 (2.20)	2.43 (2.49)	3.08 (3.13)	<0.0001	0.14 (0.03)	<0.0001
-286C/A	2.19 (2.20)	2.79 (2.63)	2.24 (2.65)	0.0096	0.17 (0.06)	0.007

Geometric mean CRP and approximate SD are shown.

*1=common allele, 2=rare allele.

The association of CRP haplotypes with CRP concentration was then evaluated. The CAT haplotype was used as a reference. CRP concentrations showed an incremental change across all four haplotypes providing some evidence for a functional variant on each haplotype (see Table 7.6).

Table 7.6. Association of CRP haplotypes with CRP concentration in the NPHSII study.

C144T/G2302A/T4899G	Mean CRP (95% CI)	P value
CAT	2.03 (1.85-2.22)	reference
CGT	2.38 (2.15-2.64)	0.04
TGT	3.05 (2.76-3.37)	<0.0001
CGG	3.60 (2.79-4.64)	<0.0005
Global p value		<0.0001

TGT v CGT: p=0.002

CGT v CGG: p=0.003

7.4.3 Balanced distribution of a variety of covariables by genotype and haplotype class in the NPHSII study

On the expectation of the random allocation of genotype and haplotype, the distribution of a range of measures among differing genotype and haplotype classes was also examined. The analysis showed that none of the 4 common CRP SNPs examined exhibited association with any of these parameters despite strong and robust associations with CRP concentration. The apparent association of the +2302G/A polymorphism with smoking (p=0.007) is likely to have arisen by chance, as the association was not particularly extreme and it was not reproduced in the Ely data set.

When covariables were examined by haplotype using the CAT haplotype as the reference, there was again no association with any risk factors for coronary disease except for a small statistically significant association with smoking (p=0.02), which is again, likely due to chance (see Table 7.11).

Table 7.7. Baseline characteristics by +1444C/T genotype in the NPHSII study (those with CRP>20 have been excluded).

	CC N=1077	CT N=914	TT N=230	P value	Change per allele** (SE)	P value
Age (years)	56.1 (3.4)	55.8 (3.4)	56.0 (3.3)	0.29	-0.12 (0.11)	0.29
BMI (kg/m^2)	26.3 (3.5)	26.4 (3.4)	26.6 (3.3)	0.53	0.11 (0.11)	0.32
Smoking	28.0 (302)	25.1 (229)	26.1 (60)	0.32	-0.087 (0.073)	0.23
Alcohol Median units/wk [IQR]	7 [2-16]	6 [2-16]	5 [1-13]	0.09	-0.097 (0.081)	0.23
SBP* (mmHg)	137.1 (18.7)	136.1 (18.2)	136.9 (17.8)	0.45	-0.003 (0.004)	0.46
Cholesterol (mmol/L)	5.72 (1.00)	5.73 (1.00)	5.79 (1.02)	0.64	0.03 (0.03)	0.42
ApoB* (mg/dL)	0.86 (0.25)	0.85 (0.23)	0.87 (0.23)	0.58	-0.003 (0.096)	0.76
ApoA (mg/dL)	1.65 (0.32)	1.66 (0.31)	1.63 (0.33)	0.48	-0.006 (0.011)	0.61
Apob/A*	0.53 (0.21)	0.52 (0.19)	0.54 (0.21)	0.36	0.0004 (0.013)	0.98
HDL* (mmol/L)	0.81 (0.23)	0.81 (0.25)	0.79 (0.24)	0.64	-0.010 (0.011)	0.38
Triglyceride* (mmol/L)	1.76 (0.93)	1.77 (0.93)	1.75 (0.91)	0.96	-0.001 (0.017)	0.97
Fibrinogen* (g/L)	2.67 (0.47)	2.66 (0.49)	2.68 (0.46)	0.74	-0.001 (0.006)	0.84

*Geometric mean(approximate sd)

**For log-normal data, change is on a log scale. For categorical variables, change is in log odds.

Table 7.8. Baseline characteristics by +2302G/A genotype in the NPHSII study (those with CRP>20 have been excluded).

	GG N=983	GA N=1009	AA N=259	P value	Change per allele** (se)	P value
Age (years)	56.0 (3.5)	55.9 (3.4)	56.4 (3.4)	0.15	0.12 (0.11)	0.28
BMI (kg/m^2)	26.4 (3.4)	26.4 (3.4)	26.1 (3.6)	0.39	-0.08 (0.11)	0.44
Smoking	23.5 (231)	27.6 (278)	32.4 (84)	0.007	0.22 (0.07)	0.002
Alcohol						
% drinkers	81.1 (797)	80.9 (816)	82.2 (213)	0.88	0.02 (0.08)	0.79
Median units/wk [IQR]	6 [2-15]	6 [2-16]	8 [2-17]	0.31		
SBP* (mmHg)	137.0 (18.2)	136.4 (18.4)	137.7 (19.6)	0.54	0.0004 (0.0042)	0.92
Cholesterol (mmol/L)	5.73 (1.00)	5.74 (0.99)	5.74 (0.98)	0.98	0.005 (0.031)	0.87
ApoB* (mg/dL)	0.86 (0.23)	0.87 (0.24)	0.85 (0.25)	0.60	0.002 (0.009)	0.79
ApoA (mg/dL)	1.64 (0.32)	1.65 (0.31)	1.66 (0.34)	0.80	0.007 (0.011)	0.51
Apob/A*	0.53 (0.20)	0.53 (0.20)	0.53 (0.22)	0.85	-0.002 (0.013)	0.88
HDL* (mmol/L)	0.81 (0.25)	0.79 (0.25)	0.83 (0.24)	0.14	-0.002 (0.011)	0.86
Triglyceride* (mmol/L)	1.75 (0.92)	1.79 (0.94)	1.73 (0.91)	0.60	0.003 (0.017)	0.87
Fibrinogen* (g/L)	2.65 (0.47)	2.69 (0.49)	2.68 (0.44)	0.19	0.008 (0.006)	0.15

*Geometric mean(approximate sd)

**For log-normal data, change is on a log scale. For categorical variables, change is in log odds.

Table 7.9. Baseline characteristics by +4899T/G genotype in the NPHSII study (those with CRP>20 have been excluded).

	TT N=2031	TG N=216	GG N=8	P value	Change per allele** (se)	P value
Age (years)	56.0 (3.4)	55.8 (3.3)	55.5 (4.6)	0.67	-0.20 (0.23)	0.37
BMI (kg/m^2)	26.3 (3.4)	26.4 (3.7)	24.8 (2.7)	0.42	-0.02 (0.23)	0.93
Smoking	26.5 (538)	25.5 (55)	37.5 (3)	0.66	-0.007 (0.152)	0.96
Alcohol						
% drinkers	81.3 (1651)	80.6 (174)	75.0 (6)	0.75	-0.068	0.68
Median units/wk [IQR]	6 [2-16]	8 [2-16]	3 [0.5-5.5]	0.26	(0.168)	
SBP* (mmHg)	136.8 (18.4)	136.6 (18.9)	135.4 (14.9)	0.96	-0.002 (0.009)	0.80
Cholesterol (mmol/L)	5.74 (0.99)	5.69 (1.04)	5.71 (1.09)	0.80	-0.04 (0.07)	0.52
ApoB* (mg/dL)	0.86 (0.24)	0.87 (0.25)	0.83 (0.23)	0.75	0.01 (0.02)	0.56
ApoA (mg/dL)	1.65 (0.31)	1.64 (0.34)	1.64 (0.20)	0.96	-0.006 (0.023)	0.78
Apob/A*	0.53 (0.20)	0.54 (0.23)	0.51 (0.14)	0.75	0.017 (0.028)	0.55
HDL* (mmol/L)	0.81 (0.25)	0.78 (0.23)	0.88 (0.29)	0.26	-0.03 (0.02)	0.28
Triglyceride* (mmol/L)	1.76 (0.92)	1.82 (1.01)	1.57 (0.78)	0.57	0.021 (0.035)	0.56
Fibrinogen* (g/L)	2.67 (0.48)	2.65 (0.46)	2.77 (0.58)	0.74	-0.003 (0.012)	0.83

*Geometric mean(approximate sd)

**For log-normal data, change is on a log scale. For categorical variables, change is in log odds.

Table 7.10. Baseline characteristics by –286C/T/A genotype in the NPHSII study (those with CRP>20 have been excluded).

	CC N=771	CT N=862	TT N=196	CA N=151	TA N=102	AA N=16	Change per T allele** (se)	P value	Change per A allele** (se)	P value
Age (years)	56.0 (3.5)	55.8 (3.4)	55.8 (3.2)	56.0 (3.3)	55.3 (3.4)	56.1 (4.3)	-0.19 (0.12)	0.11	-0.14 (0.21)	0.50
BMI (kg/m^2)	26.3 (3.4)	26.3 (3.5)	26.5 (3.5)	26.4 (3.5)	26.8 (3.5)	25.9 (3.1)	0.07 (0.12)	0.54	0.16 (0.21)	0.44
Smoking	26.5 (204)	26.3 (227)	29.1 (57)	27.2 (41)	24.5 (25)	31.3 (5)	0.028 (0.078)	0.71	0.001 (0.14)	0.94
Alcohol										
% drinkers	80.9 (624)	80.7 (696)	79.6 (156)	81.5 (123)	82.4 (84)	75.0 (12)	-0.025 (0.087)	0.77	0.01 (0.16)	0.95
Median units/wk [IQR]	6 (1-16)	6.5 (2-16)	4 (1-12)	7 (2-15)	8 (2-17)	4 (0.5-8)				
SBP* (mmHg)	137.4 (19.2)	136.3 (17.8)	137.0 (16.9)	137.2 (17.6)	135.6 (19.2)	138.6 (18.5)	-0.004 (0.005)	0.33	-0.001 (0.008)	0.86
Cholesterol (mmol/L)	5.75 (1.00)	5.73 (1.00)	5.79 (1.03)	5.67 (1.10)	5.79 (1.12)	6.05 (0.93)	0.009 (0.035)	0.79	0.018 (0.06)	0.77
ApoB* (mg/dL)	0.87 (0.25)	0.86 (0.23)	0.87 (0.22)	0.87 (0.27)	0.85 (0.24)	0.96 (0.24)	-0.006 (0.010)	0.59	0.011 (0.018)	0.55
ApoA (mg/dL)	1.66 (0.32)	1.66 (0.32)	1.64 (0.32)	1.66 (0.35)	1.63 (0.33)	1.64 (0.26)	-0.008 (0.012)	0.51	-0.011 (0.021)	0.62
Apob/A*	0.53 (0.21)	0.52 (0.20)	0.54 (0.20)	0.53 (0.24)	0.53 (0.22)	0.60 (0.21)	-0.001 (0.014)	0.93	0.018 (0.026)	0.47
HDL (mmol/L)*	0.82 (0.24)	0.80 (0.26)	0.78 (0.23)	0.78 (0.23)	0.82 (0.27)	0.86 (0.20)	-0.018 (0.0125)	0.14	-0.013 (0.022)	0.55
Triglyceride* (mmol/L)	1.78 (0.96)	1.79 (0.92)	1.79 (0.92)	1.82 (1.03)	1.76 (0.96)	1.59 (0.73)	0.003 (0.018)	0.87	-0.006 (0.032)	0.85
Fibrinogen* (g/L)	2.66 (0.46)	2.66 (0.49)	2.70 (0.49)	2.66 (0.40)	2.66 (0.51)	2.69 (0.6)	0.004 (0.006)	0.53	-0.00002 (0.011)	0.999

*Geometric mean(approximate sd). **For log-normal data, change is on a log scale. For categorical variables, change is in log odds.

Table 7.11. Baseline characteristics by haplotype in those subjects with CRP measures in the NPHSII study.

	CRP haplotypes (C1444T/G2302A/T4899G)				
	CAT 121	**CGT 111**	**TGT 211**	**CGG 112**	**P value**
Age (years)	56.0 [55.8 – 56.4]	56.2 [55.8 – 56.4]	55.8 [55.4 - 56.0]	55.6 [54.6 – 56.4]	0.30
BMI (kg/m^2)	26.2 [26.0 – 26.6]	26.2 [26.0 – 26.6]	26.4 [26.2 – 26.8]	26.2 [25.4 – 27.2]	0.84
Smoking OR	1.00	0.78 (0.66-0.92)	0.81 (0.68-0.95)	0.89 (0.65-1.21)	0.02
SBP (mmHg)	135.4 [133.5–137.3]	137.0 [133.5–140.6]	135.6 [132.2–139.2]	136.7 [126.5–147.8]	0.85
Alcohol OR	1.00	1.04 (0.85-1.28)	0.94 (0.78-1.13)	0.89 (0.63-1.25)	0.66
Cholesterol (mmol/L)	5.74 [5.64 – 5.84]	5.72 [5.42 – 5.82]	5.74 [5.66 – 5.84]	5.64 [5.40 – 5.88]	0.88
HDL (mmol/L)	0.81 [0.78 – 0.84]	0.83 [0.80 – 0.86]	0.79 [0.76 – 0.82]	0.76 [0.70 – 0.84]	0.17
ApoA (mg/dL)	1.666 [1.634–1.696]	1.642 [1.608–1.676]	1.646 [1.612–1.678]	1.628 [1.542–1.714]	0.75
ApoB (mg/dL)	0.86 [0.84 – 0.88]	0.86 [0.84 – 0.88]	0.85 [0.83 – 0.88]	0.88 [0.82 – 0.95]	0.90
ApoB/A ratio	0.53 [0.51 – 0.54]	0.53 [0.51 – 0.56]	0.53 [0.51 – 0.55]	0.55 [0.50 – 0.61]	0.83
Triglyceride (mmol/L)	1.76 [1.68 – 1.86]	1.75 [1.66 – 1.84]	1.76 [1.67 – 1.85]	1.80 [1.57 – 2.05]	0.98
Fibrinogen (g/L)	2.80 [2.79 - 2.81]	2.79 [2.78 - 2.80]	2.79 [2.78 - 2.80]	2.80 [2.77 – 2.82]	0.58
Framingham score	13.08 [12.90–13.28]	12.78 [12.58–13.00]	12.74 [12.54–12.94]	12.66 [12.20–13.10]	0.08
Homocysteine (μmol/L)	12.30 [11.82–12.81]	11.73 [11.27–12.21]	12.33 [11.85–12.81]	12.83 [11.59–14.21]	0.25
Lp(a) OR	1.00	0.95 (0.82-1.11)	1.13 (0.97-1.32)	0.94 (0.70-1.26)	0.16

Results are based on an additive model and two copies of the haplotype showing mean (95% CI) values, except for smoking, alcohol and Lp(a) where results are odds ratios. For Lp(a), the results are odds of being above median.

7.4.4 Influence of age and other covariables on the genotype/haplotype-CRP association in the NPHSII study

To test whether the association of genotype and haplotype with CRP was modified by other variables relevant to cardiovascular disease risk, the data was examined after dividing participants into quartiles of the respective covariate. Genotype/haplotype and other covariables each influenced CRP concentration, but in proportionate terms, the effect of genotype/haplotype on CRP concentration was preserved across all strata (see Table 7.12). The combined effects of haplotype and covariables were in line with those expected from individual effects of SNPs, with no evidence for interaction (see Figures 7.1-7.4).

For the –286C/T/A polymorphism, the high-CRP genotypes (TT, TA, CA and AA) were grouped together and compared with the CC and CT genotypes. CRP concentrations increased across the genotypes for each quartile of covariate examined. Conversely, CRP concentrations increased in a graded manner across increasing quartiles of the covariables examined in all genotypic classes. There did not appear to be any interaction between any of the covariates studied and CRP –286C/T/A genotype (see Table 7.13 and Figures 7.5-7.8).

In other words, CRP genotype and haplotype appeared to exert a similar proportional effect on CRP concentration at all levels of age, blood pressure, BMI, cholesterol, triglycerides and fibrinogen examined.

Table 7.12. Estimated geometric mean (95% CI) for CRP by haplotype and strata of covariates in the NPHSII study.

	CRP haplotypes (C1444T/G2302A/T4899G)					Interaction
Quartile of variable	**CAT**	**CGT**	**TGT**	**CGG**	**P value**	**p value**
Age						
1: ≤54	1.94 (1.64-2.30)	2.31 (1.91-2.80)	2.69 (2.25-3.21)	2.33 (1.55-3.52)	0.12	0.37
2: 55-57	1.92 (1.58-2.33)	2.31 (1.89-2.81)	3.22 (2.67-3.89)	3.90 (2.01-7.52)	0.002	
3: 58-59	1.72 (1.40-2.11)	2.32 (1.76-3.06)	2.61 (2.06-3.31)	4.80 (2.87-8.00)	0.009	
4: ≥60	2.57 (2.18-3.03)	2.53 (2.14-2.99)	4.05 (3.29-4.98)	5.99 (3.13-11.45)	0.0002	
BMI						
1: <24.1	1.31 (1.07-1.60)	1.73 (1.36-2.22)	1.93 (1.57-2.37)	2.37 (1.33-4.23)	0.05	0.96
2: 24.1-26.1	1.94 (1.63-2.32)	2.19 (1.80-2.69)	2.86 (2.32-3.52)	3.00 (1.81-4.98)	0.06	
3: 26.2-28.4	2.51 (1.51-2.96)	2.62 (2.22-3.10)	3.39 (2.81-4.09)	3.68 (2.25-6.00)	0.09	
4: >28.4	2.80 (2.38-3.30)	3.27 (2.77-3.87)	4.30 (3.65-5.07)	6.36 (4.20-9.62)	0.0002	
SBP						
1: ≤122	1.51 (1.25-1.82)	2.04 (1.66-2.52)	2.42 (1.97-2.97)	4.31 (2.60-7.11)	0.0005	0.02
2: 122.5-134	1.90 (1.56-2.30)	2.43 (1.96-3.00)	2.54 (2.06-3.13)	1.76 (1.01-3.05)	0.19	
3: 134.5-147	2.53 (2.14-3.00)	2.44 (2.04-2.92)	3.39 (2.82-8.26)	2.93 (1.63-5.28)	0.08	
4: ≥147	2.43 (2.05-2.87)	2.62 (2.16-3.16)	4.14 (3.46-4.95)	7.16 (4.15-12.35)	<0.0001	
Cholesterol						
1: ≤5	1.78 (1.46-2.16)	1.88 (1.53-2.32)	2.62 (2.12-3.23)	2.25 (1.34-3.77)	0.08	0.54
2: 5.1-5.6	1.77 (1.48-2.12)	2.32 (1.87-2.86)	2.64 (2.13-3.27)	4.26 (2.38-7.64)	0.008	
3: 5.6-6.3	2.18 (1.82-2.60)	2.82 (2.33-3.40)	3.01 (2.54-3.57)	3.60 (2.19-5.91)	0.06	
4: ≥6.3	2.46 (2.10-2.87)	2.68 (2.21-3.24)	4.30 (3.54-5.23)	5.81 (3.50-9.68)	<0.0001	
HDL						
1: <0.65	2.59 (2.10-3.17)	2.93 (2.29-3.74)	3.49 (2.81-4.33)	3.19 (2.05-4.97)	0.33	0.48
2: 0.65-0.81	2.33 (1.90-2.85)	2.78 (2.19-3.52)	2.57 (2.01-3.29)	3.22 (1.98-5.25)	0.63	
3: 0.81-0.99	1.87 (1.51-2.31)	1.88 (1.48-2.39)	2.90 (2.32-3.63)	5.32 (2.27-12.48)	0.002	
4: >0.99	1.46 (1.17-1.83)	1.73 (1.35-2.21)	2.15 (1.66-2.79)	1.90 (0.92-3.93)	0.28	
ApoB						
1: <0.73	1.59 (1.30-1.95)	1.92 (1.50-2.46)	2.49 (1.98-3.13)	2.64 (1.26-5.57)	0.05	0.76
2: 0.73-0.85	1.78 (1.45-2.19)	2.74 (2.18-3.44)	2.93 (2.37-3.62)	3.15 (1.99-4.98)	0.009	
3: 0.85-1.0	2.29 (1.89-2.76)	2.60 (2.11-3.19)	3.67 (3.02-4.47)	3.91 (2.18-7.02)	0.01	
4: >1.0	2.73 (2.26-3.29)	2.47 (2.03-3.00)	3.82 (3.09-4.72)	4.46 (2.70-7.34)	0.02	
ApoA						
1: ≤1.44	2.27 (1.85-2.78)	2.93 (2.39-3.58)	3.89 (3.16-4.79)	3.35 (2.10-5.33)	0.01	0.67
2: 1.45-1.64	2.54 (2.10-3.07)	2.42 (1.94-3.01)	3.18 (2.57-3.93)	3.31 (1.77-6.20)	0.28	
3: 1.65-1.85	2.08 (1.71-2.51)	2.12 (1.70-2.65)	3.23 (2.58-4.07)	4.44 (2.52-7.81)	0.005	
4: ≥1.85	1.53 (1.25-1.86)	2.17 (1.70-2.78)	2.44 (1.99-3.00)	3.07 (1.68-5.62)	0.01	
ApoB/ApoA						
1: <0.41	1.44 (1.18-1.74)	1.82 (1.41-2.33)	2.40 (1.93-2.99)	2.87 (1.44-5.72)	0.007	
2: 0.41-0.52	1.96 (1.58-2.43)	2.64 (2.08-3.34)	2.75 (2.24-3.37)	3.51 (2.18-5.65)	0.08	0.26
3: 0.52-0.65	2.85 (2.39-3.41)	2.18 (1.79-2.64)	3.45 (2.80-4.26)	2.83 (1.73-4.63)	0.03	
4: >0.65	2.34 (1.92-2.84)	3.03 (2.49-3.70)	4.26 (3.42-5.30)	4.72 (2.78-8.03)	0.001	
Triglyceride						
1: <1.22	1.42 (1.17-1.73)	1.78 (1.43-2.22)	2.28 (1.86-2.80)	3.27 (1.81-5.93)	0.006	0.92
2: 1.22-1.71	1.62 (1.34-1.97)	2.01 (1.64-2.46)	2.88 (2.33-3.55)	2.93 (1.84-4.68)	0.001	
3: 1.71-2.46	2.44 (2.04-2.93)	2.76 (2.29-3.33)	3.21 (2.69-3.83)	6.61 (2.31-5.51)	0.19	
4: >2.46	3.00 (2.61-3.45)	3.24 (2.72-3.86)	4.18 (3.48-5.00)	5.43 (3.35-8.80)	0.02	
Fibrinogen						
1: <238	1.05 (0.89-1.24)	1.46 (1.21-1.75)	2.13 (1.76-2.58)	2.30 (1.42-3.74)	<0.0001	0.10
2: 238-265	1.63 (1.37-1.95)	2.11 (1.76-2.52)	2.66 (2.20-3.21)	3.44 (2.21-5.35)	0.001	
3: 266-298	2.42 (2.05-2.86)	2.62 (2.15-3.20)	3.46 (2.89-4.15)	4.72 (2.77-8.05)	0.02	
4: >298	3.89 (3.24-4.67)	4.56 (3.72-5.61)	4.30 (3.66-5.04)	4.10 (2.78-6.03)	0.75	

Means are for 2 copies of the haplotype.

Figure 7.1. Estimated geometric mean for CRP by haplotype and quartiles of BMI in NPHSII.

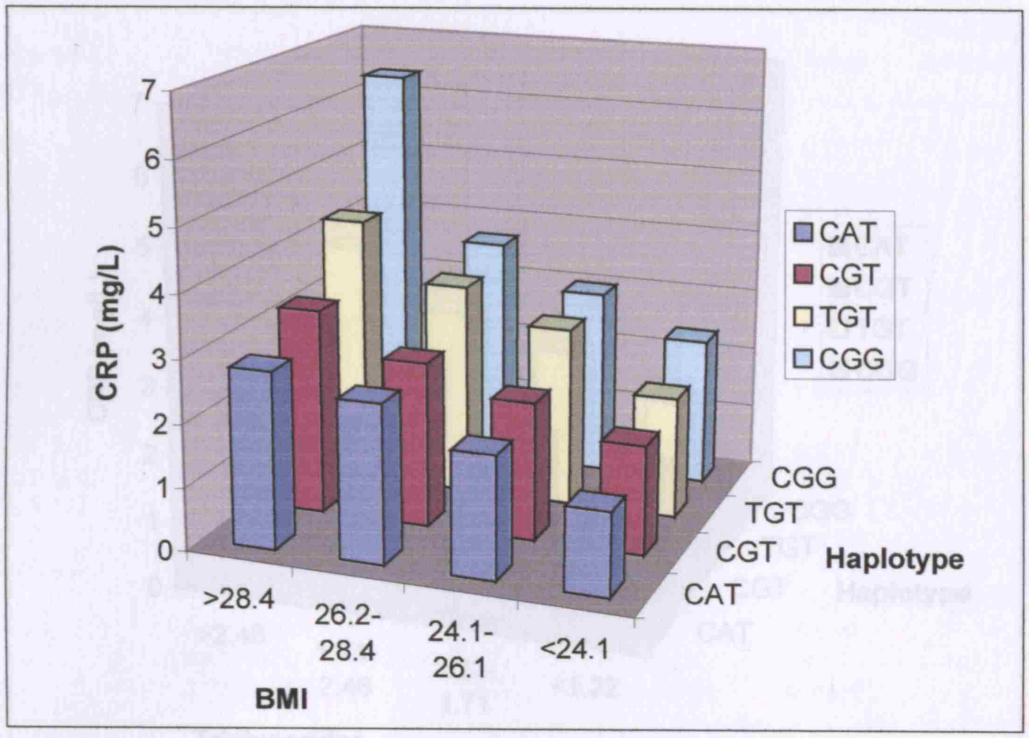


Figure 7.2. Estimated geometric mean for CRP by haplotype and quartiles of systolic blood pressure showing significant interaction in NPHSII.

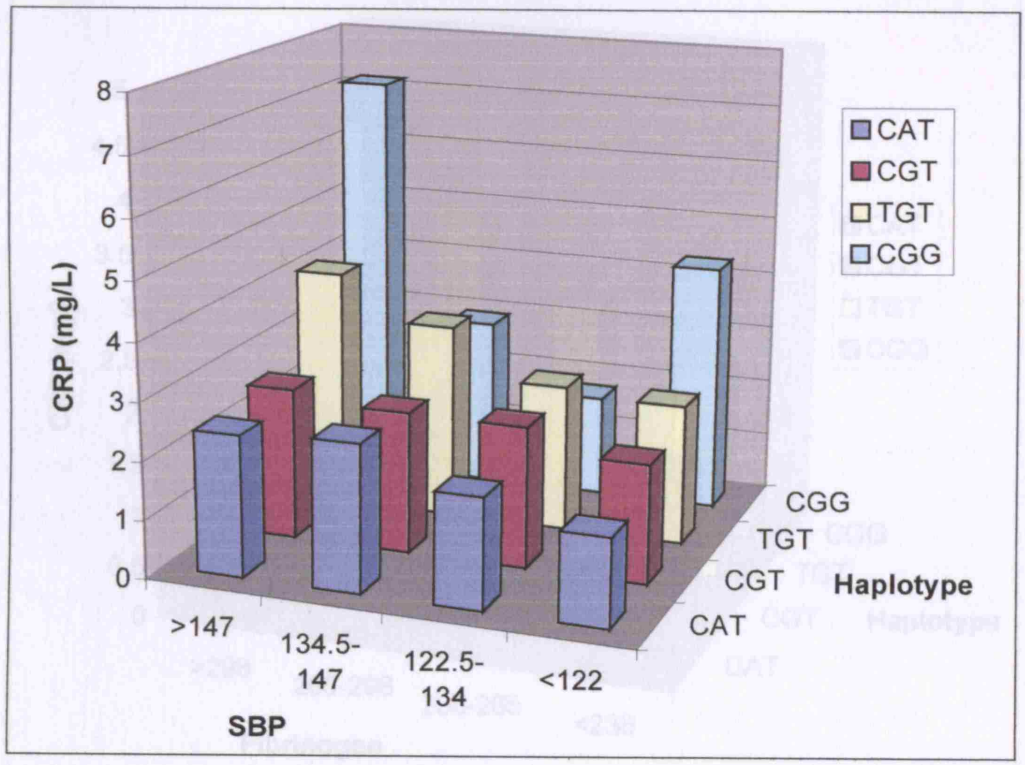


Figure 7.3. Estimated geometric mean for CRP by haplotype and quartiles of triglycerides in NPHSII.

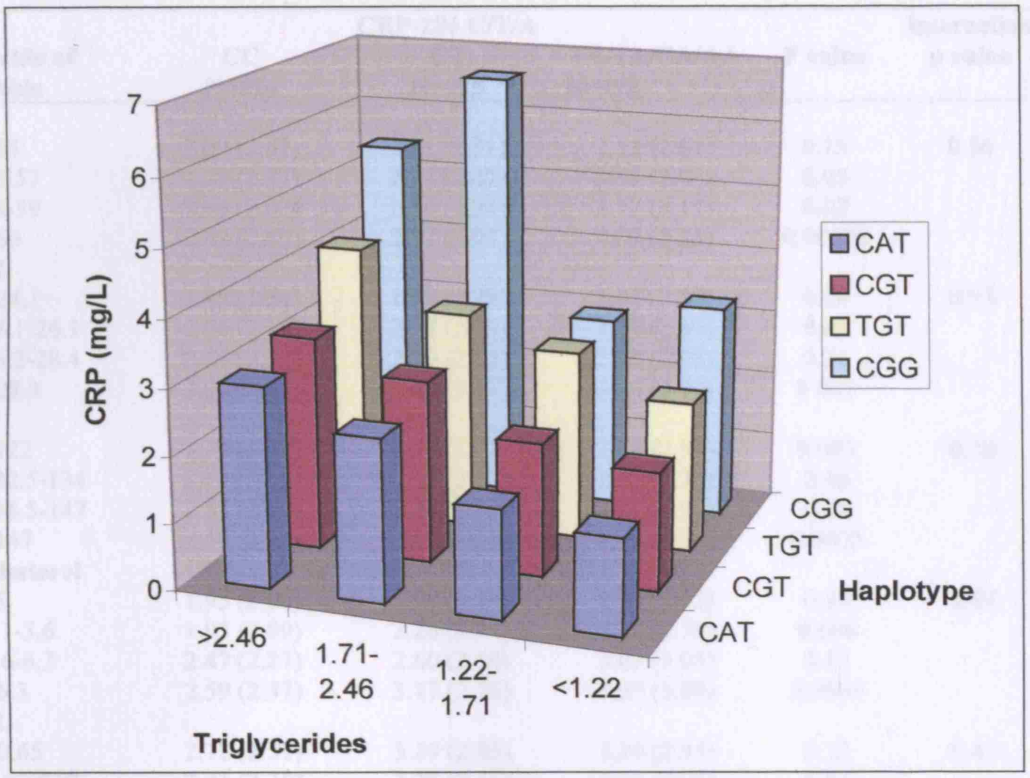


Figure 7.4. Estimated geometric mean for CRP by haplotype and quartiles of fibrinogen in NPHSII.

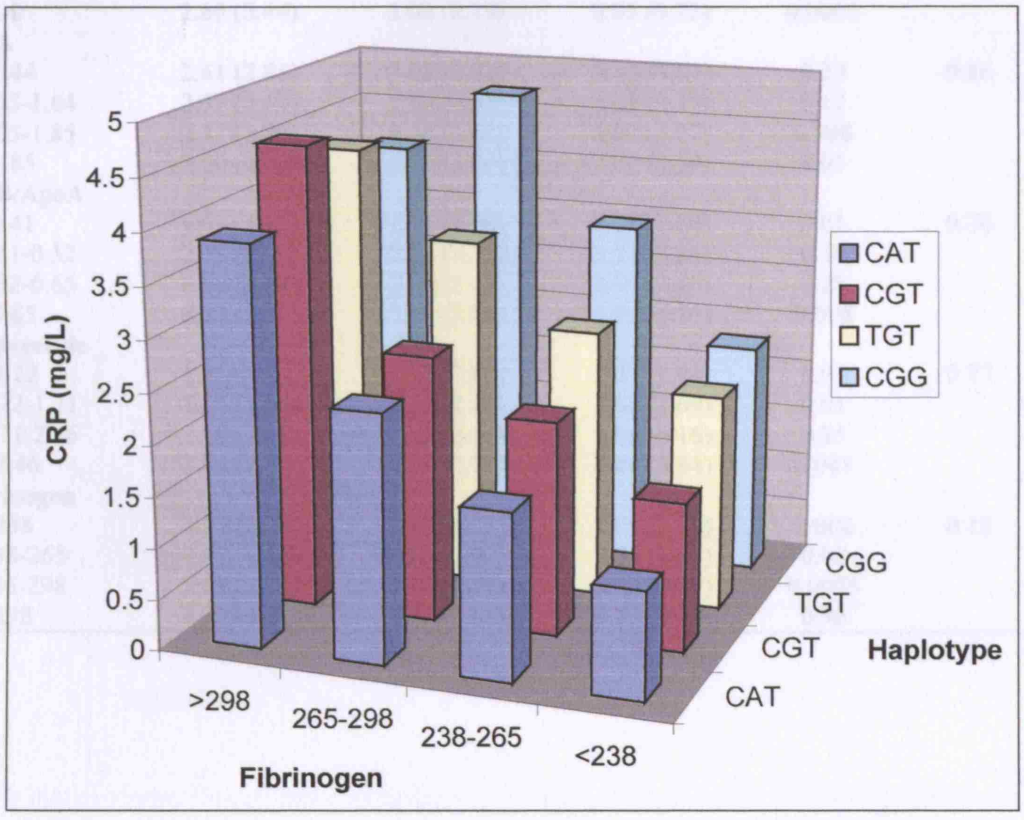


Table 7.13. Geometric mean (approximate SD) for CRP by –286C/T/A genotype and strata of covariates.

Quartile of variable	CRP-286 C/T/A CC N=824	CT N=925	TT/TA/CA/AA N=492	P value	Interaction p value
Age					
1: ≤54	2.09 (2.07)	2.31 (2.31)	2.53 (2.64)	0.15	0.56
2: 55-57	2.28 (2.37)	2.35 (2.47)	2.98 (2.84)	0.05	
3: 58-59	2.00 (2.07)	2.30 (2.34)	2.77 (3.17)	0.07	
4: ≥60	2.49 (2.40)	2.97 (3.01)	4.09 (3.56)	0.00008	
BMI					
1: <24.1	1.49 (1.54)	1.59 (1.79)	1.97 (2.28)	0.09	0.93
2: 24.1-26.1	2.06 (2.14)	2.29 (2.22)	2.79 (2.69)	0.03	
3: 26.2-28.4	2.66 (2.38)	2.79 (2.64)	3.20 (2.98)	0.21	
4: >28.4	3.00 (2.65)	3.61 (3.05)	4.26 (3.70)	0.001	
SBP					
1: ≤122	1.76 (1.83)	1.94 (1.93)	2.64 (2.98)	0.003	0.36
2: 122.5-134	2.13 (2.19)	2.13 (2.27)	2.42 (2.30)	0.46	
3: 134.5-147	2.51 (2.36)	2.71 (2.52)	3.07 (2.98)	0.18	
4: ≥147	2.59 (2.49)	3.32 (3.36)	3.93 (3.79)	0.0005	
Cholesterol					
1: ≤5	1.95 (2.08)	1.99 (2.24)	2.33 (2.52)	0.29	0.67
2: 5.1-5.6	1.95 (2.09)	2.26 (2.34)	2.86 (2.70)	0.006	
3: 5.6-6.3	2.47 (2.27)	2.60 (2.54)	3.07 (3.05)	0.12	
4: ≥6.3	2.59 (2.37)	3.17 (2.76)	3.97 (3.89)	0.0003	
HDL					
1: <0.65	2.72 (2.51)	3.19 (2.80)	3.10 (2.95)	0.32	0.45
2: 0.65-0.81	2.41 (2.39)	2.28 (2.36)	3.14 (2.95)	0.04	
3: 0.81-0.99	1.97 (2.04)	2.28 (2.33)	2.81 (2.63)	0.04	
4: >0.99	1.62 (1.64)	1.80 (1.84)	1.85 (2.37)	0.58	
ApoB					
1: <0.73	1.79 (1.80)	1.85 (2.14)	2.58 (2.71)	0.02	0.31
2: 0.73-0.85	2.16 (2.18)	2.62 (2.56)	2.52 (2.71)	0.18	
3: 0.85-1.0	2.46 (2.29)	2.82 (2.79)	3.03 (3.20)	0.19	
4: >1.0	2.60 (2.44)	3.02 (2.79)	4.05 (3.73)	0.0005	
ApoA					
1: ≤1.44	2.61 (2.85)	3.08 (2.92)	3.18 (3.61)	0.20	0.84
2: 1.45-1.64	2.55 (2.47)	2.69 (2.62)	3.22 (3.29)	0.17	
3: 1.65-1.85	2.11 (1.95)	2.36 (2.52)	3.09 (3.07)	0.008	
4: ≥1.85	1.80 (1.60)	2.02 (2.24)	2.52 (2.50)	0.03	
ApoB/ApoA					
1: <0.41	1.67 (1.55)	1.75 (1.98)	2.38 (2.60)	0.01	0.70
2: 0.41-0.52	2.16 (2.04)	2.52 (2.73)	2.72 (2.74)	0.16	
3: 0.52-0.65	2.55 (2.39)	2.89 (2.47)	2.99 (2.80)	0.29	
4: >0.65	2.72 (2.79)	3.10 (3.02)	3.94 (4.16)	0.008	
Triglyceride					
1: <1.22	1.58 (1.77)	1.78 (2.05)	2.27 (2.44)	0.01	0.77
2: 1.22-1.71	1.86 (1.80)	2.10 (2.26)	2.68 (2.69)	0.01	
3: 1.71-2.46	2.68 (2.31)	2.72 (2.51)	3.08 (3.16)	0.35	
4: >2.46	3.06 (2.76)	3.51 (2.73)	4.40 (3.64)	0.001	
Fibrinogen					
1: <238	1.32 (1.29)	1.47 (1.44)	1.95 (1.82)	0.002	0.13
2: 238-265	1.91 (1.68)	2.15 (2.02)	2.54 (2.37)	0.02	
3: 266-298	2.38 (2.16)	2.81 (2.74)	3.63 (3.52)	0.0005	
4: >298	4.22 (3.71)	4.32 (3.81)	4.23 (4.49)	0.96	

Figure 7.5. Estimated geometric mean for CRP by –286C/T/A genotype and quartiles of BMI in NPHSII.

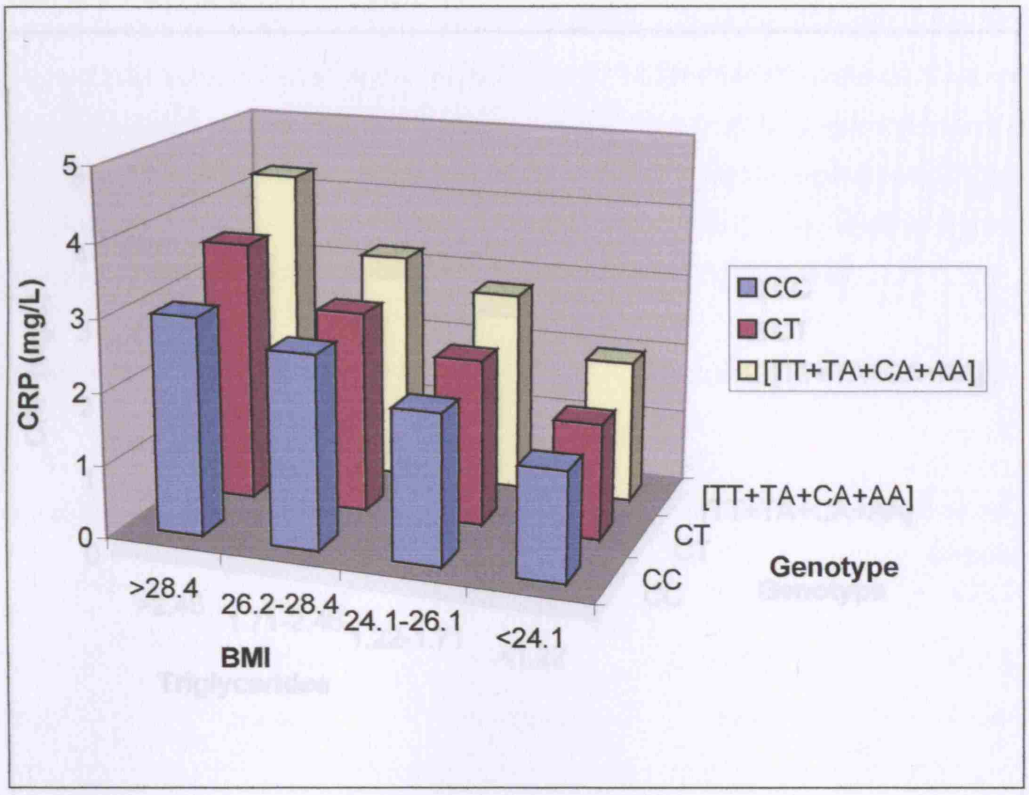


Figure 7.6. Estimated geometric mean for CRP by –286C/T/A genotype and quartiles of systolic blood pressure in NPHSII.

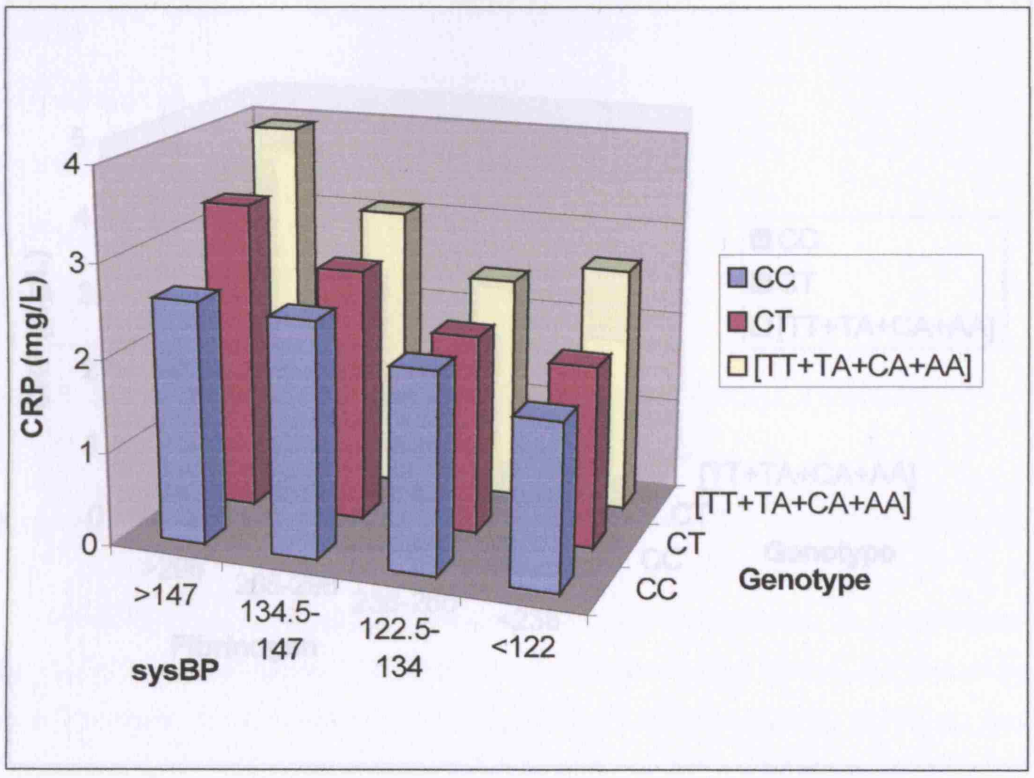


Figure 7.7. Estimated geometric mean for CRP by –286C/T/A genotype and quartiles of triglycerides in NPHSII.

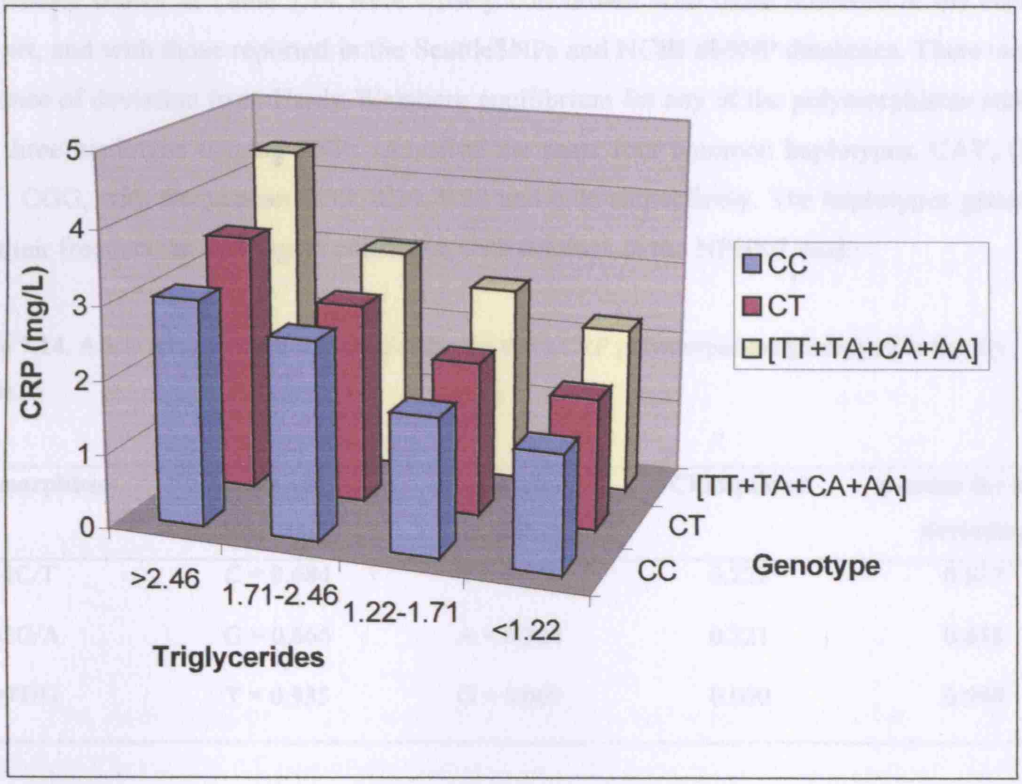
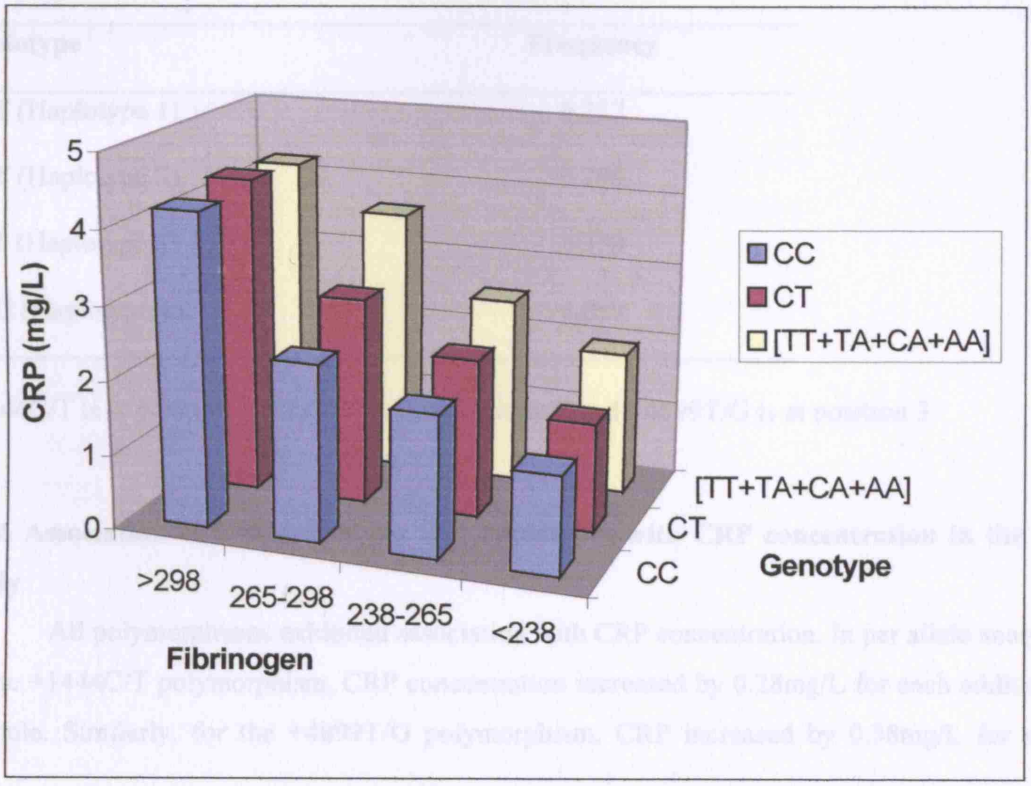


Figure 7.8. Estimated geometric mean for CRP by –286C/T/A genotype and quartiles of fibrinogen.



7.4.5 CRP genotypes and haplotypes in the Ely study

The same three tagging SNPs in the CRP gene were typed in the Ely cohort. Allele frequencies shown in Table 7.14 were closely concordant with those observed in the NPHSII data set, and with those reported in the SeattleSNPs and NCBI dbSNP databases. There was no evidence of deviation from Hardy-Weinberg equilibrium for any of the polymorphisms studied. The three haplotype tagging SNPs identified the same four common haplotypes, CAT, CGT, TGT, CGG, with frequencies 0.33, 0.29, 0.32 and 0.06 respectively. The haplotypes generated and their frequencies were again consistent with findings in the NPHSII study.

Table 7.14. Allele and genotype frequencies for the three CRP polymorphisms genotyped in the Ely cohort.

Polymorphism	Major allele frequency	Minor allele frequency	Chi squared	p-value for H-W deviation
+1444C/T	C = 0.684	T = 0.316	0.222	0.637
+2302G/A	G = 0.666	A = 0.334	0.221	0.638
+4899T/G	T = 0.935	G = 0.065	0.000	0.999

Table 7.15. Haplotypes generated from the three tagging SNPS (+1444C/T, +2302G/A and +4899T/G) with a frequency of 5% or more in the Ely cohort.

Haplotype	Frequency
CAT (Haplotype 1)	0.327
CGT (Haplotype 2)	0.293
TGT (Haplotype 3)	0.320
CGG (Haplotype 4)	0.059

+1444C/T is at position 1, +2302G/A is at position 2 and +4899T/G is at position 3.

7.4.6 Association of CRP genotypes and haplotypes with CRP concentration in the Ely study

All polymorphisms exhibited association with CRP concentration. In per allele analyses of the +1444C/T polymorphism, CRP concentration increased by 0.28mg/L for each additional T-allele. Similarly, for the +4899T/G polymorphism, CRP increased by 0.38mg/L for each

additional G-allele. In per allele analyses of the +2302G/A polymorphism, CRP concentration decreased by 0.37mg/L for each additional A-allele (see Table 7.16).

Table 7.16. Association of CRP genotypes with CRP concentration (mg/L) in the Ely study (those with CRP>20mg/L have been excluded).

SNP	Genotype*			P-value	Change per allele (SE)	P-value
	11	12	22			
+1444C/T	0.91 (1.00)	1.18 (1.28)	1.66 (1.61)	0.0002	0.28 (0.07)	<0.0001
+2302G/A	1.33 (1.40)	1.01 (1.11)	0.57 (0.53)	<0.0001	-0.37 (0.07)	<0.0001
+4899T/G	1.02 (1.08)	1.48 (1.77)	2.65 (7.57)	0.01	0.38 (0.13)	0.003

Geometric mean CRP and approximate SD are shown.

*1=common allele, 2=rare allele.

The three tagging SNPs were then used in combination as before to generate haplotypes. The same CAT haplotype was used as a reference and the association of haplotype with CRP concentrations was examined. CRP concentrations increased across the haplotypic classes as seen in the NPHSII study, although absolute values of CRP were lower in this data set (see Table 7.17).

Table 7.17. Association of CRP haplotypes with CRP concentration in the Ely study.

C144T/G2302A/T4899G	Mean CRP (95% CI)	P value
CAT	0.69 (0.56-0.86)	reference
CGT	1.56 (1.25-1.93)	<0.0001
TGT	1.06 (0.83-1.36)	0.011
CGG	1.98 (1.27-3.11)	<0.0001
Global p value		<0.0001

CGT v TGT: p=0.01

CGT v CGG: p=0.028

7.4.7 Balanced distribution of a variety of covariables by genotype and haplotype class in the Ely study

Since the allocation of genotype and haplotype is expected to be a random process, the distribution of a range of measures among differing genotype and haplotype classes was also examined on the expectation that the distributions of these variables would be balanced. The analysis showed that none of the CRP SNPs examined exhibited association with any of the covariates examined, despite their strong and robust associations with CRP concentration, with the exception of the +1444C/T polymorphism, which showed a small statistically significant association with BMI (p=0.02). Again, this is likely to reflect the play of chance as the effect size is small, the p-value is not particularly extreme and the effect was not seen in the NPHSII cohort. When baseline measures were examined by haplotype using the CAT haplotype as the reference, there was no association with any risk factors for coronary disease (see Table 7.21).

Table 7.18. Baseline characteristics by +1444C/T genotype in the Ely study (those with CRP>20 have been excluded).

	CC N=259	CT N=232	TT N=57	P value	Change per allele** (SE)	P value
Age (years)	53.6 (8.0)	54.0 (7.7)	52.5 (7.0)	0.40	-0.21 (0.50)	0.68
BMI* (kg/m^2)	25.2 (3.8)	25.6 (4.0)	26.6 (3.9)	0.06	0.02 (0.01)	0.02
Smoking OR	18.3 (47)	16.0 (37)	24.6 (14)	0.32	0.08 (0.17)	0.62
Alcohol OR	73.4 (190)	74.6 (173)	77.2 (44)	0.83	0.09 (0.15)	0.55
SBP* (mmHg)	128.2 (16.8)	127.8 (16.8)	130.2 (14.7)	0.63	0.004 (0.008)	0.64
Cholesterol (mmol/L)	6.46 (1.29)	6.35 (1.32)	6.70 (1.31)	0.19	0.006 (0.013)	0.66
HDL* (mmol/L)	1.44 (0.37)	1.43 (0.36)	1.42 (0.38)	0.84	-0.009 (0.017)	0.58
Triglyceride* (mmol/L)	1.22 (0.65)	1.17 (0.58)	1.33 (0.61)	0.24	0.01 (0.03)	0.71

*Geometric mean (approximate SD).

**For log-normal data change is on log scale. For categorical variables change is in log odds.

Table 7.19. Baseline characteristics by +2302G/A genotype in the Ely study (those with CRP>20 have been excluded).

	GG N=263	GA N=271	AA N=64	P value	Change per allele** (SE)	P value
Age (years)	53.6 (7.8)	54.2 (7.7)	53.1 (7.5)	0.50	0.04 (0.48)	0.94
BMI* (kg/m^2)	26.0 (3.9)	25.1 (3.6)	25.5 (3.9)	0.02	-0.02 (0.01)	0.05
Smoking OR	18.3 (48)	16.4 (44)	17.2 (11)	0.84	-0.07 (0.17)	0.66
Alcohol OR	74.1 (195)	74.5 (202)	73.4 (47)	0.98	-0.005 (0.142)	0.97
SBP* (mmHg)	128.4 (16.0)	128.1 (17.6)	129.0 (14.5)	0.93	0.001 (0.008)	0.94
Cholesterol (mmol/L)	6.42 (1.30)	6.46 (1.39)	6.52 (1.06)	0.84	0.007 (0.013)	0.56
HDL* (mmol/L)	1.41 (0.37)	1.43 (0.34)	1.48 (0.40)	0.30	0.024 (0.016)	0.13
Triglyceride* (mmol/L)	1.22 (0.60)	1.19 (0.60)	1.17 (0.64)	0.70	-0.026 (0.031)	0.41

*Geometric mean (approximate SD).

**For log-normal data change is on log scale. For categorical variables change is in log odds.

Table 7.20. Baseline characteristics by +4899T/G genotype in the Ely study (those with CRP>20 have been excluded).

	TT N=516	TG N=70	GG N=2	P value	Change per allele** (SE)	P value
Age (years)	53.9 (7.7)	53.7 (7.9)	48.5 (2.2)	0.61	-0.42 (0.93)	0.66
BMI* (kg/m^2)	25.5 (3.7)	25.7 (4.2)	23.8 (3.9)	0.67	0.006 (0.018)	0.73
Smoking OR	17.7 (91)	15.7 (11)	50.0 (1)	0.41	-0.001 (0.32)	0.998
Alcohol OR	75.4 (389)	71.4 (50)	100.0 (2)	0.70	-0.11 (0.27)	0.68
SBP* (mmHg)	128.1 (16.5)	129.7 (16.4)	122.0 (2.4)	0.66	0.008 (0.016)	0.61
Cholesterol (mmol/L)	6.47 (1.34)	6.16 (1.09)	6.45 (1.62)	0.15	-0.04 (0.02)	0.07
HDL* (mmol/L)	1.43 (0.36)	1.40 (0.33)	1.73(0.35)	0.44	-0.008 (0.031)	0.79
Triglyceride* (mmol/L)	1.21 (0.62)	1.12 (0.50)	1.34 (0.76)	0.40	-0.07 (0.060)	0.25

*Geometric mean (approximate SD).

**For log-normal data change is on log scale. For categorical variables change is in log odds.

Table 7.21. Baseline characteristics by haplotype in those subjects with CRP measures in the Ely cohort.

	CRP haplotypes (C1444T/G2302A/T4899G)				
	CAT 121	**CGT 111**	**TGT 211**	**CGG 112**	**P value**
Frequency	0.327	0.293	0.320	0.059	
Age (years)	53.8	53.4	54.0	52.3	0.86
	(52.2-55.4)	(51.8-55.0)	(52.4-55.6)	(48.6-56.1)	
BMI* (kg/m^2)	24.9	26.3	25.3	25.9	0.09
	(24.1-25.6)	(25.5-27.1)	(24.5-26.2)	(24.1-27.8)	
Smoking OR	1.00	1.09	0.98	0.95	0.99
		(0.51-2.32)	(0.42-2.25)	(0.21-4.32)	
Alcohol OR	1.00	1.13	0.95	0.64	0.82
		(0.57-2.27)	(0.47-1.90)	(0.19-2.11)	
SBP* (mmHg)	128.5	128.8	127.2	127.7	0.94
	(125.2-131.9)	(125.5-132.3)	(124.0-130.6)	(118.9-137.0)	
Cholesterol* (mmol/L)	6.48	6.51	6.44	5.71	0.13
	(6.21-6.77)	(6.26-6.78)	(6.17-6.71)	(5.04-6.48)	
HDL* (mmol/L)	1.49	1.42	1.40	1.40	0.44
	(1.42-1.56)	(1.36-1.49)	(1.33-1.47)	(1.22-1.61)	
Triglyceride* (mmol/L)	1.49	1.23	1.24	1.04	0.64
	(1.07-1.29)	(1.11-1.37)	(1.12-1.37)	(0.78-1.40)	

Results are based on an additive model and two copies of the haplotype showing mean (95% CI) values, except for smoking and alcohol where results are odds ratios.

*Log transformed.

7.4.8 Influence of age and other covariables on the haplotype-CRP association in the Ely study

The effect of CRP haplotypes on CRP concentration was re-evaluated among subjects categorised into quantiles of measured covariables. CRP concentrations increased across the haplotypes within each quantile of covariate examined, and increased within each haplotype with increasing quantiles of each covariate (see Table 7.22). In proportionate terms, the effect of haplotype on CRP was preserved across all strata with no evidence for interaction, with the exception of alcohol intake, which gave a marginal interaction p value of 0.03 and may simply be due to chance (see Figures 8.9-8.12).

Table 7.22. Estimated geometric mean (95% CI) for CRP by haplotype and strata of covariates.

Strata	CRP haplotypes (C1444T/G2302A/T4899G) CAT	CGT	TGT	CGG	P value	Interaction p value
Age						
1: ≤45	0.61 (0.31-1.17)	0.72 (0.40-1.28)	0.80 (0.39-1.67)	3.14 (1.02-9.68)	0.13	0.32
2: 45-50	0.45 (0.30-0.66)	0.78 (0.46-1.34)	1.86 (1.20-2.89)	0.42 (0.10-1.94)	0.0002	
3: 51-55	0.61 (0.35-1.04)	1.27 (0.64-2.51)	1.31 (0.77-2.22)	2.79 (0.84-9.30)	0.07	
4: 56-59	0.86 (0.50-1.47)	1.01 (0.57-1.80)	1.75 (1.09-2.80)	3.10 (1.68-8.90)	0.13	
5: ≥60	1.03 (0.68-8.90)	1.46 (0.98-2.18)	2.09 (1.43-3.05)	2.02 (0.91-4.45)	0.04	
BMI						
1: <24.1	0.45 (0.32-0.63)	0.73 (0.49-1.07)	1.15 (0.82-1.61)	0.86 (0.37-1.96)	0.005	0.56
2: 24.1-26.1	0.88(0.55-1.42)	1.00 (0.57-1.74)	1.02 (0.57-1.83)	2.08 (0.58-2.00)	0.69	
3: 26.2-28.4	0.72 (0.44-1.19)	1.47 (0.93-2.32)	2.05 (1.37-3.07)	1.91 (0.94-3.87)	0.01	
4: >28.4	1.36 (1.12-2.06)	2.06 (1.30-3.27)	2.31 (1.65-3.23)	6.47 (2.74-1.53)	0.02	
Smoking						
Non	0.60 (0.47-0.76)	0.96 (0.74-1.24)	1.57 (1.26-1.96)	1.70 (1.06-2.71)	<0.0001	0.25
Current	1.22 (0.72-2.08)	1.66 (0.82-3.37)	1.61 (0.86-3.02)	4.23 (1.17-15.33)	0.56	
Alcohol						
Non	0.51 (0.33-0.79)	2.06 (1.35-3.13)	1.32 (0.85-2.06)	3.81 (1.78-8.20)	0.0001	0.03
Current	0.78 (0.61-1.00)	0.96 (0.72-1.29)	1.43 (1.11-1.84)	1.47 (0.82-2.64)	0.006	
SBP						
1: ≤122	0.59 (0.42-0.83)	0.88 (0.58-1.33)	1.32 (0.92-1.88)	1.09 (0.44-2.71)	0.03	0.97
2: 122.5-134	0.62 (0.40-0.97)	1.06 (0.66-1.72)	1.58 (1.00-2.51)	3.67 (1.36-9.95)	0.008	
3: 134.5-147	0.87 (0.56-1.36)	1.20 (0.70-2.05)	1.62 (1.00-2.64)	2.42 (1.19-4.92)	0.10	
4: ≥147	0.83 (0.46-1.49)	1.65 (0.97-2.82)	2.13 (1.38-3.27)	2.10 (0.27-16.02)	0.08	
Cholesterol						
1: ≤5	0.70 (0.25-1.98)	1.13 (0.50-2.54)	1.15 (0.44-3.03)	0.44 (0.02-9.14)	0.84	0.65
2: 5.1-5.6	0.53 (0.32-0.89)	0.81 (0.39-1.68)	0.99 (0.57-1.72)	1.39 (0.60-3.26)	0.26	
3: 5.6-6.3	0.58 (0.35-0.97)	1.07 (0.5-2.03)	0.97 (0.56-1.70)	2.26 (1.08-4.70)	0.08	
4: ≥6.3	0.75 (0.58-0.98)	1.15 (0.84-1.56)	2.01 (1.53-2.65)	3.69 (1.92-7.10)	<0.0001	
HDL						
1: <1.0	0.85 (0.44-1.63)	1.09 (0.61-1.94)	2.71 (1.67-4.40)	0.28 (0.06-1.27)	0.02	0.16
2: >1.0	0.67 (0.53-0.84)	1.05 (0.80-1.37)	1.43 (1.12-1.83)	2.26 (1.41-3.63)	<0.0001	
Triglyceride						
1: <1.22	0.47 (0.35-0.63)	0.82 (0.60-1.14)	1.26 (0.94-1.68)	1.17 (0.61-2.25)	<0.0001	0.52
2: 1.22-1.71	1.28 (0.82-2.02)	1.16 (0.70-1.91)	1.57 (1.02-2.43)	3.85 (1.80-8.23)	0.25	
3: 1.71-2.46	1.01 (0.64-1.59)	1.47 (0.92-2.37)	2.14 (1.36-3.38)	3.52 (0.84-14.70)	0.01	
4: >2.46	1.86 (0.77-4.49)	2.17 (1.14-4.12)	2.44 (2.01-3.74)	8.80 (3.23-19.7)	0.14	

Means are for 2 copies of the haplotype.

Figure 7.9. Estimated geometric mean for CRP by haplotype and BMI in the Ely study.

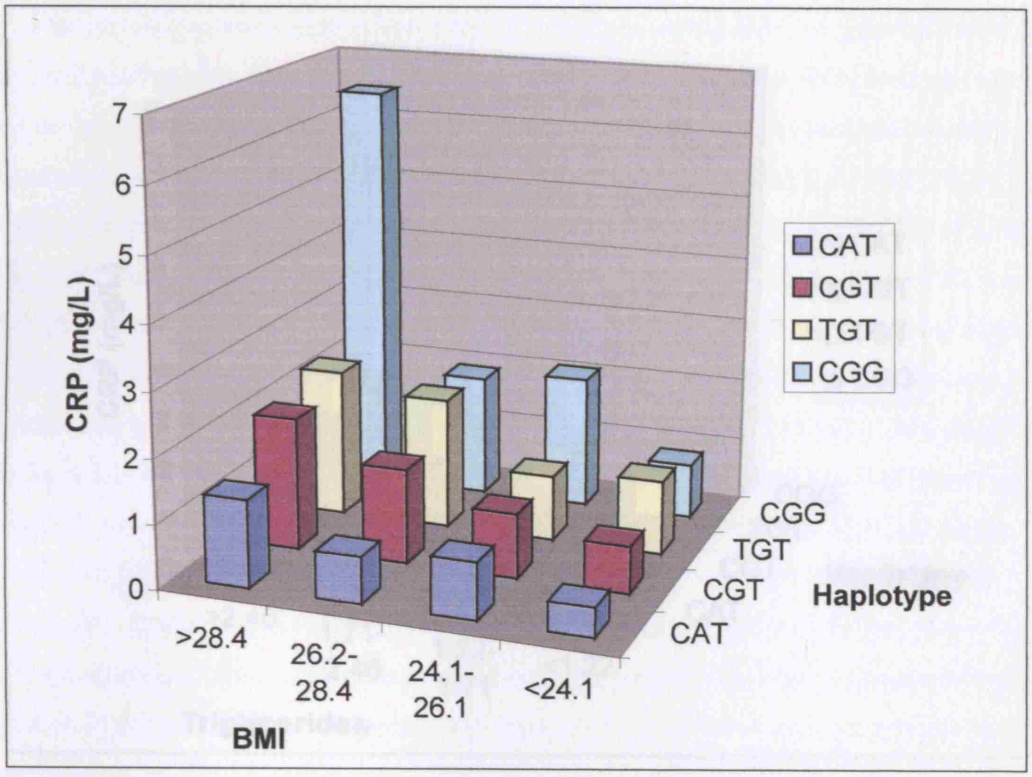


Figure 7.10. Estimated geometric mean for CRP by haplotype and systolic blood pressure in the Ely study.

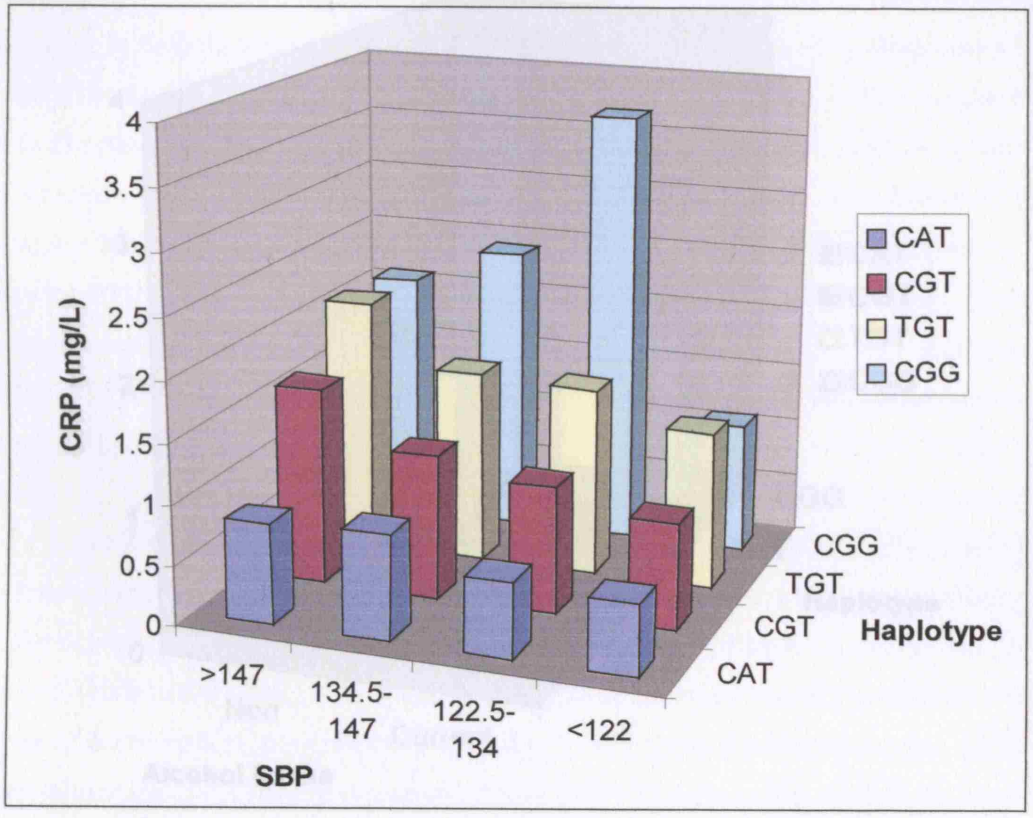


Figure 7.11. Estimated geometric mean for CRP by haplotype and triglycerides in the Ely study.

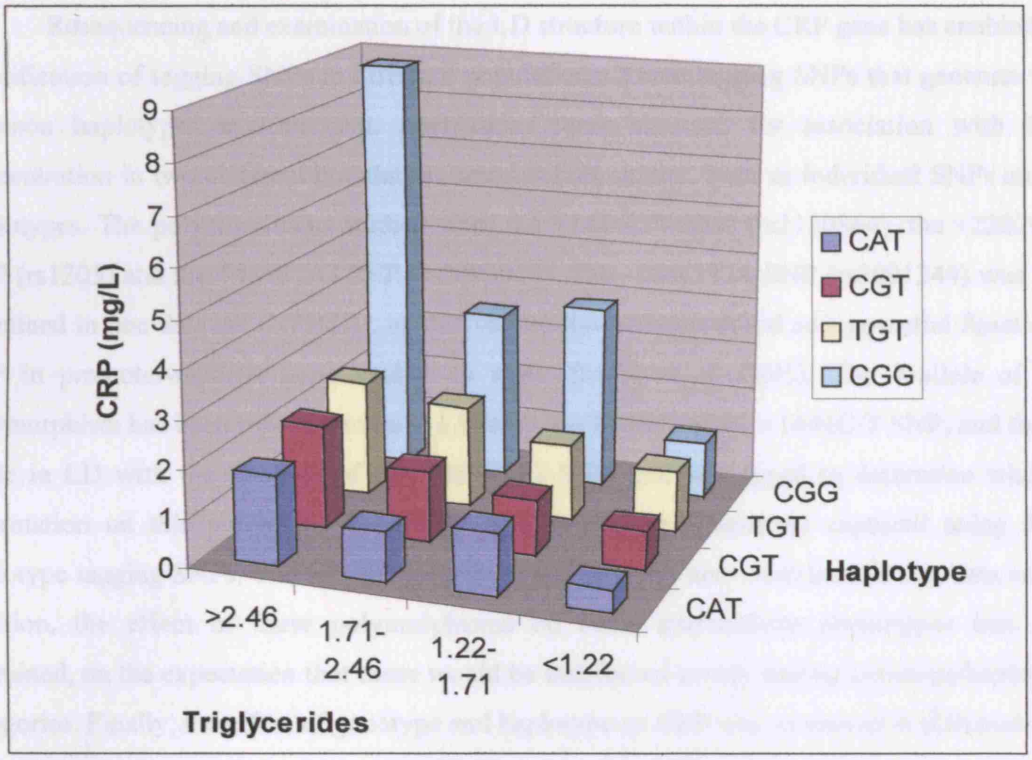
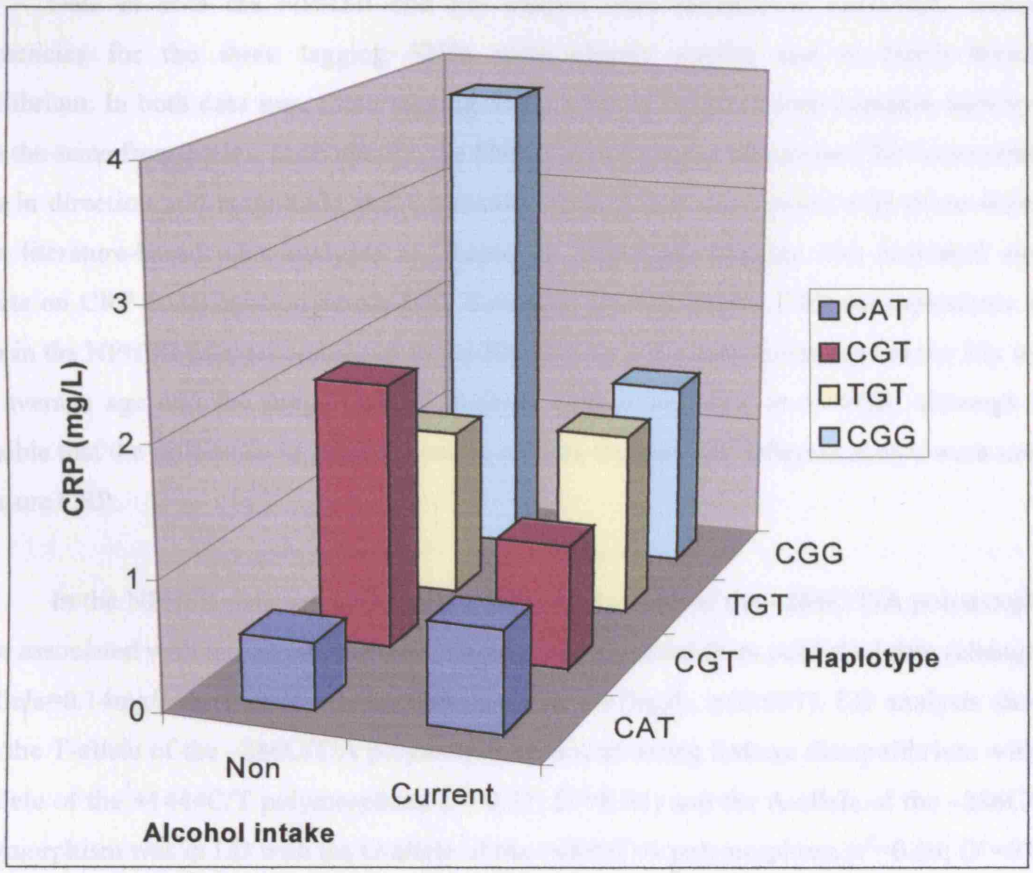


Figure 7.12. Estimated geometric mean for CRP by haplotype and alcohol intake in the Ely study.



7.5 Discussion

Resequencing and examination of the LD structure within the CRP gene has enabled the identification of tagging SNPs in different populations. Three tagging SNPs that generate four common haplotypes in Caucasian populations were assessed for association with CRP concentration in two different population-based cohort studies, both as individual SNPs and as haplotypes. The polymorphisms studied were the +1444C/T SNP (rs1130864), the +2302G/A SNP (rs1205) and the +4899T/G SNP (rs3093077). The –286C/T/A SNP (rs3091244) was also examined in one data set (NPHSII), as this variant has been proposed as a potential functional SNP in promoter-reporter gene studies *in vitro* (Szalai *et al.* 2005). The T-allele of this polymorphism has been reported to be in LD with the T-allele of the +1444C/T SNP, and the A-allele in LD with the G-allele of the +4899T/G SNP, and was typed to determine whether information on this potentially functional SNP would be adequately captured using three haplotype tagging SNPs. The LD between these variants was also examined in this data set. In addition, the effect of these polymorphisms on other intermediate phenotypes was also examined, on the expectation that these would be distributed evenly among genotype/haplotype categories. Finally, the effect of genotype and haplotype on CRP was examined at different ages and at different levels of blood pressure, BMI, cholesterol, triglycerides and other covariables in stratified analyses.

Data in both the NPHSII and Ely studies were remarkably consistent. Genotype frequencies for the three tagging SNPs were closely similar and in Hardy-Weinberg equilibrium. In both data sets, these tagging SNPs inferred the same four common haplotypes, with the same frequencies. Individually, the SNPs showed similar effects on CRP concentration, both in direction and magnitude and the results were closely concordant with those obtained from literature-based meta-analyses in Chapter 6. Haplotype analyses also indicated similar effects on CRP concentration across both data sets. Overall, higher CRP concentrations were seen in the NPHSII data set compared to the Ely data set. This may be because in the Ely study, the average age and the proportion of smokers were lower than in NPHSII, although it is possible that the difference in absolute values reflects the fact that different assays were used to measure CRP.

In the NPHSII data set, both the T-allele and A-allele of the –286C/T/A polymorphism were associated with increased CRP concentrations as expected from published data (change per T-allele=0.14mg/L, $p<0.0001$; change per A-allele=0.17mg/L, $p\leq0.007$). LD analysis showed that the T-allele of the –286C/T/A polymorphism was in strong linkage disequilibrium with the T-allele of the +1444C/T polymorphism ($r^2=0.77$; D'=0.91) and the A-allele of the –286C/T/A polymorphism was in LD with the G-allele of the +4899T/G polymorphism ($r^2=0.69$; D'=0.94).

Data on other intermediate phenotypes including age, BMI, systolic blood pressure, cholesterol and triglycerides were also available in both data sets. In the NPHSII data set, no associations were seen between CRP genotypes and these covariates, except for smoking, which was significantly associated with the +2302G/A polymorphism (p=0.002). Similarly, in the Ely data set, only one significant association was seen between BMI and the +1444C/T polymorphism (p=0.02). When data on baseline covariates by haplotype were examined, there was no significant difference across the haplotypes in either data set, except for a small significant association with smoking in the NPHSII data set (p=0.02). These two significant associations may therefore simply represent false positive associations arising due to chance, because of the multiple numbers of comparison tests undertaken. This suggests that the effect of CRP genotype and haplotype is not confounded by these variables that are strongly associated with CRP itself as was shown in Chapter 4.

Apart from investigating the relationship between genotype and CRP concentration, it was also possible to take advantage of the varying range of phenotypes measured, to test whether the genotype/haplotype-CRP association was modified by measures associated with vascular disease. The effects of genotype and haplotype on CRP concentration were preserved across all strata of covariables examined, with a similar proportionate effect on CRP concentration at all levels of age, blood pressure, BMI and so on. In the NPHSII data set, when interaction between covariate and haplotype was examined, the CRP concentration differences were as expected from their independent main effects, with no evidence for interaction, with the exception of systolic blood pressure (p=0.02). Similarly, in the Ely data set, there was no evidence for significant interaction with the covariates studied except for smoking (p=0.03). However, these two significant interactions may be due to multiple testing resulting in false positives by chance. These results suggest that there is no systematic difference in the proportionate effect of genotype or haplotype on CRP among subjects with more extreme levels of age, blood pressure, BMI and other variables that themselves influence CRP.

Published data is now available on haplotype and CRP concentration in three large studies. The study by Carlson et al. in the Coronary Artery Risk Development in Young Adults (CARDIA) population-based cohort used seven tagging SNPs to generate haplotypes as the cohort was of mixed ancestry, and in contrast to Caucasian subjects, individuals of African ancestry exhibit ten common haplotypes at this locus (Carlson *et al.* 2005). Nevertheless, the haplotypes that contained the rare alleles of the –286C/T/A, +1444C/T and +4899T/G polymorphisms were all associated with increased CRP concentrations as in the NPHSII and Ely studies. The haplotype containing the rare allele of the +2302G/A polymorphism was associated with lower CRP compared to the haplotypes containing the common allele, again

consistent with the findings in this chapter. Similar differences in CRP concentration were seen across haplotypes generated from three tagging SNPs in the British Women's Heart and Health study and in the Rotterdam study, where the haplotype containing the rare allele of the +1444C/T polymorphism was associated with the highest CRP concentration and the haplotype containing the rare allele of the +2302G/A SNP was associated with the lowest CRP concentration (Timpson *et al.* 2005; Kardys *et al.* 2006). Therefore, the haplotype-CRP associations seen in this chapter are highly concordant with published data.

The work in this chapter shows that CRP genotypes and haplotypes are robustly associated with CRP concentration, similar to the associations seen between β-fibrinogen promoter variants and fibrinogen levels (Davey Smith *et al.* 2005b), APOAV polymorphisms and triglyceride levels (Talmud *et al.* 2002; Baum *et al.* 2003), or haplotypes in the cholesteryl ester transfer protein (CETP) gene and HDL-cholesterol levels (Thompson *et al.* 2005). Genotype-intermediate phenotype associations are now being readily identified and much more consistently than genotype-disease associations. This is likely to be because (i) the difference in a continuous phenotype is a more immediate consequence of genetic variation than the disease outcome; (ii) these phenotypes are readily measured traits that are less prone to misclassification than disease outcomes, and (iii) are continuous rather than disease outcomes (e.g. disease/no disease) and so the power to detect such associations is increased in comparison to studies where the outcome is categorical.

Where genotypes or haplotypes are reliably associated with intermediate phenotypes of interest (as is the case with CRP), studying genotype/haplotype associations with disease could reduce confounding and reverse causality bias seen in observational studies where the intermediate phenotype is measured directly, and so may allow greater understanding of the causal relevance of these intermediate phenotypes. The analyses done in this chapter confirm that there is no confounding by measured covariates in the genotype/haplotype-CRP association. They also illustrate that the effect on the gene is preserved at all levels of age, blood pressure, BMI and other covariables that influence CRP. Therefore, these polymorphisms appear to be suitable for use in a Mendelian randomisation analysis to understand the causal relevance of CRP in coronary disease or other outcomes. These analyses are the subject of further work described in Chapter 10.

7.6 Conclusions

Genotyping in two population-based cohort studies was carried out to obtain precise estimates of the effect of individual CRP SNPs and haplotypes on CRP concentration. The minor alleles of the –286C/T/A, +1444C/T and +4899T/G polymorphisms were all associated with higher CRP concentration in the NPHSII and Ely data sets. The minor allele of the +2302G/A polymorphism was associated with lower CRP concentration. These results are consistent with the associations with CRP concentration seen in published data. The effect of genotype on CRP concentration did not appear to be confounded by other variables also associated with CRP concentration, even when the data were stratified by quantiles of covariates. The +1444C/T, +2302G/A and +4899T/G polymorphisms were used to generate four common CRP haplotypes, which were also associated with CRP concentration in both data sets. Baseline covariates did not appear to be associated with haplotype, overall and when the data were stratified, suggesting the association between haplotype and CRP concentration is not confounded by these covariates. The causal relevance of CRP for coronary events may be better understood by using CRP genotype or haplotype as a non-confounded proxy for CRP itself.

8. Effect of genotype on CRP levels following inflammatory stimuli

8.1 Aim

To investigate the association of CRP genotypes (and haplotypes) and CRP concentration during an acute inflammatory response.

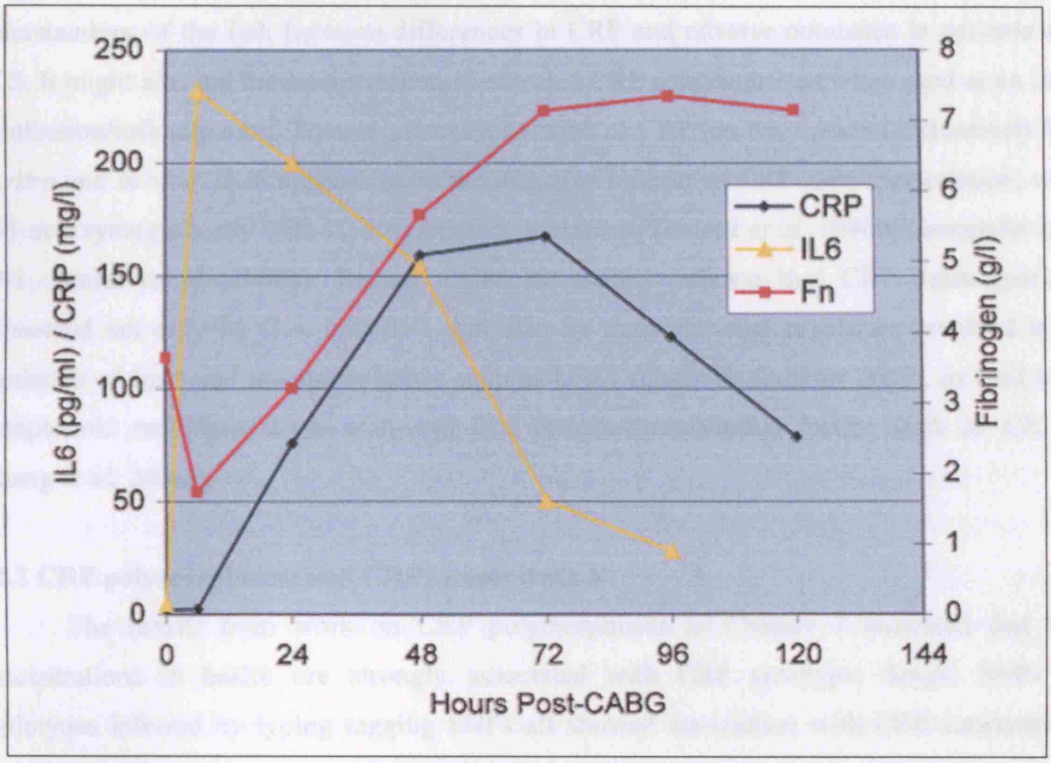
8.2 Background

8.2.1 CRP in inflammation

The acute phase reactant, C-reactive protein, was first identified in the 1930s as a circulating protein that binds the C-polysaccharide on *S. pneumoniae* (Tillet & Francis 1930). Further studies showed that it was synthesised by the liver. CRP circulates at a concentration in the range of 0.1-5mg/L in healthy individuals (Pepys *et al.* 1979). However, concentrations can increase up to 1000mg/L within 24-48 hours following an infective or inflammatory episode. This substantial increase in CRP concentration is sustained if the stimulus is ongoing (Pepys & Hirschfield 2003). Early studies examining the acute phase response found that the initial peak in CRP concentration following an inflammatory stimulus occurs after the more rapid alterations in inflammatory cytokines such as IL-1β and IL-6 (Fransen *et al.* 2000; Fulop *et al.* 2001; Craig *et al.* 2001). A recent study in subjects undergoing coronary artery bypass surgery showed an early peak in IL-6 concentration a few hours after surgery, followed by a peak in CRP concentration at around 48 hours, with a more gradual and later rise in fibrinogen that peaked at around 92 hours after surgery that was sustained until discharge from hospital (Brull *et al.* 2003) (see Figure 8.1).

Since, the hepatic clearance of CRP occurs at a constant rate, giving CRP a half-life in the circulation of 19 hours, differences in concentration during inflammation are due mainly to changes in the rate of synthesis of CRP and a major part of the regulation of its synthesis is at the level of gene transcription (Pepys & Hirschfield 2003).

Figure 8.1. Acute phase response for IL-6, CRP and fibrinogen following coronary artery bypass surgery (Brull *et al.* 2003).



Fn refers to fibrinogen.

Because CRP provides such a good index of the status of an infective or inflammatory episode, it is widely used in clinical practice as a marker of progress and response to treatment in infective or inflammatory disorders such as pneumonia, meningitis and some autoimmune diseases (Povoa *et al.* 2005; Almirall *et al.* 2004; Paradowski *et al.* 1995; Yildirim *et al.* 2004).

More recently, attention has focussed on the observation that CRP concentration increases soon after an acute coronary syndrome (ACS), and that the degree of this elevation is associated with recurrent cardiovascular events in the short-term (Berk *et al.* 1990; Haverkate *et al.* 1997; Morrow *et al.* 1998; Biasucci *et al.* 1999b; Biasucci *et al.* 1999a; Biasucci *et al.* 1999d; Pepys & Hirschfield 2001). A recent study, with a large sample size of over 7000 patients with ACS found higher concentrations of CRP at presentation were associated with a higher mortality rate within 30 days of hospital admission (odds ratio 1.19 per increasing quartile of CRP, p=0.006) (James *et al.* 2003). The results from these studies suggest that the measurement of CRP might have utility in risk stratification following ACS beyond that provided by current risk scores such as the TIMI score, or sensitive assays of myocardial injury such as troponin assays (Foussas *et al.* 2005). If so, this might help to identify patients who would benefit from more intensive therapies including coronary revascularisation by angioplasty and stent insertion.

Understanding the factors influencing the acute fluctuations in CRP concentration during infection and inflammation, and the molecular mechanisms involved may help the understanding of the link between differences in CRP and adverse outcomes in patients with ACS. It might also aid the interpretation of elevated CRP concentrations when used as an index of infection/inflammation. Transcriptional regulation of CRP has been studied extensively both *in vitro* and *in vivo*. IL-6 appears to be the principal inducer of CRP gene transcription, while IL-1 acts synergistically with IL-6 to enhance its effect (Toniatti *et al.* 1990b; Ganapathi *et al.* 1991; Szalai *et al.* 2000a). Recent molecular studies indicate that CRP transcription is influenced not only by IL-6 and IL-1, but also by transcriptional regulators involved in the regulation of lipid and metabolic genes such as USF1 (Corre & Galibert 2005), as well as by endoplasmic reticulum stress pathways that include transcription factors such as CREBH (Zhang *et al.* 2006a).

8.2.2 CRP polymorphisms and CRP concentration

The results from work on CRP polymorphisms in Chapter 7 indicated that CRP concentrations in health are strongly associated with CRP genotype. Single SNPs and haplotypes inferred by typing tagging SNPs all showed association with CRP concentration with differences in the range of 0.68-0.89mg/L being noted between homozygous individuals. For many of the SNPs, effects were shown to be consistent in a systematic review and meta-analysis of published data involving several thousand subjects (see Chapter 6). However, little is known about the effect of these variants on CRP concentration during inflammation or infection.

Previous studies of the +1444C/T polymorphism showed that healthy subjects with the T-allele exhibited a higher peak CRP concentration following 48 hours of military exercise, and that patients with the T-allele also exhibited a higher peak CRP after coronary artery bypass graft (CABG) surgery (Brull *et al.* 2003). The same association of genotype with peak CRP concentration was seen in patients undergoing intensive local treatment for severe periodontal disease, a procedure that induces a transient systemic inflammatory response (D'Aiuto *et al.* 2005).

However, since these studies, the CRP gene has been resequenced, and additional gene variants have been identified (see Chapter 5). Linkage disequilibrium across the gene has also been examined and has allowed tagging SNPs to be identified, (which can capture nearly all the genetic variation at this locus) and have been used more frequently to examine genetic associations in recent years in acknowledgement that single SNP associations might provide an incomplete assessment of genetic effects. Experimental studies *in vitro* have also identified

potential functional variants in the CRP gene promoter, such as the –286C/T/A polymorphism that lies within an E-box transcription factor binding site for the transcription factor USF1. The aim of the current study was therefore to investigate the effects of CRP genotypes and haplotypes on the peak CRP response during an acute inflammatory episode.

8.2.3 Choice of inflammatory model to study CRP

Studying CRP during naturally occurring infectious or inflammatory episodes is difficult because these occur unpredictably. However, previous studies have shown individuals with periodontal disease have a low-grade chronic inflammatory response (Lowe 2001). Moreover, when the disease is actively treated by instrumentation and scaling, changes in cytokines and acute phase reactants can be measured in the circulation. Since periodontal treatment is performed routinely at the Eastman Dental Hospital, UCL, subjects receiving treatment for this disorder are readily available for study. The time course of changes following periodontal treatment in inflammatory factors including CRP, IL-1β, IL-6 and TNF-α have been well documented, and the changes are closely similar to those that follow naturally occurring clinical infection with bacteria such as *Chlamydia pneumoniae* (Gattone *et al.* 2001; Verkerk *et al.* 2003), or after infusion of lipopolysaccharide (LPS) (Warren & Chedid 1987; Fiuza & Suffredini 2001). Therefore, the treatment of periodontal disease by instrumentation provides a useful model for studying the effects of CRP genotype and haplotype on CRP concentration during acute inflammation.

8.3 Methods

8.3.1 Study Subjects

This study was conducted on healthy individuals of mixed ethnicity referred to the Eastman Dental Hospital, UCL, UK. Only subjects presenting with severe and generalized periodontitis were invited to participate in the study. Exclusion criteria included known systemic diseases, history and/or presence of other acute or chronic infections, systemic antibiotic treatment in the preceding three months, treatment with any medication known to affect the serum level of inflammatory markers and pregnant or lactating females. A complete medical history and standard clinical periodontal parameters were collected at baseline by clinical staff. Subjects thereafter received either an intensive session of subgingival mechanical instrumentation under local anaesthesia or standard periodontal treatment of scraping by clinical staff at the department of Periodontology at the Eastman Dental Hospital. Serial blood samples were obtained at baseline, one, seven and thirty days after periodontal therapy. An original set of 55 participants was recruited for this study on which intensive periodontal treatment was carried out (data published), following which another 134 patients were additionally recruited for a randomised clinical trial where half the patients received the intensive periodontal

treatment and the other half received standard treatment. Only patients receiving the intensive treatment were included in the analysis. In addition, 45 patients were later recruited who also received intensive periodontal treatment and fulfilled the inclusion criteria for the study.

8.3.2 Measurement of CRP

Serum CRP concentrations were assessed by an automated immunoturbidimetric high-sensitivity assay (Cobas Integra, Roche AG Diagnostics, Mannheim, Germany) with a detection limit of 0.25 mg/L and inter-assay, and intra-assay coefficient of variation of 4% and 5%, respectively. Laboratory measurements were carried out in a blinded fashion and in single batches to limit inter-assays variability.

8.3.3 Genotyping the CRP polymorphisms

DNA was extracted by means of the QIAamp DNA blood minikit (QIAGEN©, Hilden, Germany) from dipotassium EDTA anticoagulated blood. The CRP +1444C/T polymorphism was genotyped by PCR and RFLP analysis using the restriction enzyme *SduI* in the original study of 55 samples, and by TaqMan assay (Applied Biosystems) in the remaining samples using the forward primer 5' GGT CTG GGA GCT CGT TAA CTA TG 3' and the reverse primer 5' TCC AAC TTG AAA AAC AAA ACA CCT CAA 3' (see Methods chapter). All other polymorphisms were also genotyped using a standard TaqMan assay by design, with the exception of the –286C/T/A triallelic polymorphism, which was genotyped using pyrosequencing (see Chapter 3).

The -717A/G polymorphism was genotyped with the forward primer 5' GCT GAG AAA ATG TGT CCA TGC AAA A 3' and the reverse primer 5' TCC TGT GTC CAA GTA TTC TCA TTG TTC 3'. The -305G/A polymorphism was genotyped with the forward primer 5' GGG CTG AAG TAG GTG TTG GA 3' and the reverse primer 5' TCC TGC GAA AAT AAT GGG AAA TGG T 3'. The +2302G/A polymorphism was genotyped with the forward primer 5' CAC CAG TAG CCA TCT TGT TTG C 3' and the reverse primer 5' CCA CTT CCA GTT TGG CTT CTG T 3'. The +4899T/G polymorphism was genotyped using the forward primer 5' TTA TCC TAG GAC AAC TGC CCA CTA 3' and the reverse primer 5' GGA GCT GAA GAG AAG GAA TCC A 3'.

The –286C/T/A polymorphism was genotyped using a pyrosequencing method using the forward primer 5' TGA TTT GGG CTG AAG TAG GTG 3', the reverse primer 5' TGG CTA TCT ATC CTG CGA AAA T 3' and the sequencing primer 5' ACC CAG ATG GCC ACT 3'. All DNA analysis was performed by staff blinded to the clinical status of the patients.

8.3.4 Sample size

In prior work, the absolute difference in peak CRP concentration, 24 hours after intensive periodontal therapy was approximately 10mg/L between subjects with CC and TT genotypes for the +1444C/T polymorphism (a difference of approximately 2 standard deviations) (D'Aiuto *et al.* 2005). Sample size estimates for the current study, in which the primary outcome was the difference in peak CRP by genotype, were based on this pilot data and are shown in Table 8.1. In subsidiary analyses, the difference in baseline CRP by genotype was also evaluated, with the expectation that the current study would be underpowered to detect the expected differences noted in the large-scale studies discussed in Chapter 7.

Table 8.1. Sample size estimates for the number of homozygous subjects required to detect the given differences in peak CRP (expressed in number of SDs) following periodontal treatment.

Difference in peak CRP between subjects homozygous for CRP polymorphisms (SD)	**Number of subjects required* (5% significance level)**	
	90% power	**80% power**
2.19	5	4
1.75	7	6
1.31	13	10
0.88	28	21
0.44	109	82

*For a polymorphism for which the frequency of individuals homozygous for the rare variant is 10%, 50 subjects require genotyping on average to identify 5 rare allele homozygotes.

8.3.5 Statistical analysis

These analyses were mainly conducted by Dr. Francesco D'Aiuto (Eastman Dental Institute, UCL) and Ms. Jackie Cooper. I contributed to the preparation of tabular data and helped to conduct some of the analyses.

Preliminary analysis of normality was performed using the Shapiro-Wilk test. Logarithmic transformations were used for CRP because of the skewed distribution, and data were reported as geometric mean and approximate standard deviation (SD). Changes in serum concentrations of CRP following periodontal therapy were used as the outcome variable in a one-way ANCOVA analysis adusting for potential confounders including age, gender, ethnicity, body mass index (BMI; expressed in kg/m2), smoking, blood pressure, periodontal diagnosis, number of teeth extracted and IL-6. Post hoc analyses were performed by Bonferroni corrections. A χ^2 test was used to compare genotype frequencies according to the Hardy-

Weinberg equilibrium. Data were analysed with the statistical software package SPSS (SPSS version 11, Chicago, IL).

Haplotype analysis was performed using a maximum likelihood model based on the stochastic-EM algorithm implemented in the THESIAS program (http://www.genecanvass.org). THESIAS allows the simultaneous estimation of haplotype frequencies and of their associated effects on the phenotype of interest. The haplotype combining the most frequent alleles at each locus was used as the reference. A global p value was calculated using differences in log-likelihood assuming an additive model of haplotype effects. Haplotypes with a frequency of less than 2% were excluded.

8.4 Results

8.4.1 CRP genotype and haplotype frequencies

Six SNPs were genotyped in the CRP gene. Allele frequencies are shown in Table 8.2. There was no evidence of deviation from Hardy-Weinberg equilibrium for the polymorphisms studied with the exception of the –286C/T/A variant. The deviation seen with this SNP may be due to the data set comprising a mixed group of individuals of differing ethnicity (see Chapter 9).

Table 8.2. Allele frequencies for each polymorphism.

Polymorphism	Major allele frequency	Minor allele frequency	Chi squared	p-value for H-W deviation
-717A/G	A = 0.744	G = 0.256	0.890	0.345
-305G/A	G = 0.959	A = 0.041	1.024	0.312
-286C/T/A	C = 0.578	T = 0.296	19.543	0.001
		A = 0.126		
+1444C/T	C = 0.638	T = 0.362	2.521	0.112
+2302G/A	G = 0.697	A = 0.303	1.808	0.179
+4899T/G	T = 0.885	G = 0.115	1.586	0.208

The six haplotype tagging SNPs identified ten haplotypes, referred to as AGCCGT, AGTCGT, GGCCGT, AATCGT, AGACGG, AGCCAT, AGTTGT, GGCTGT, AGATGG and AGCTAT, with reference to SNPs at the following polymorphic sites -717A/G / -305G/A / -286C/T/A / +1444C/T / +2302G/A / +4899T/G respectively, all of which had frequencies greater than 0.02 (see Table 8.3). Based on work in Chapter 5, the six tagging SNPs were

expected to generate ten common haplotypes with a frequency greater than 0.05 in African descent populations. Since the Periodontal disease subjects were of mixed ethnicity, the frequencies of some haplotypes were lower than expected due to differences in allele frequencies of individual SNPs, and therefore a cut off frequency of 0.02 was chosen. From the ten haplotypes observed, several corresponded to the common haplotypes seen in European populations. These were Haplotypes 1-4, 5, 6 and 7-8 (see Table 8.3).

Table 8.3. Haplotypes generated from the six tagging SNPS (-717A/G, -305G/A, -286C/T/A, +1444C/T, +2302G/A and +4899T/G) with a frequency of 2% or more.

Haplotype	Frequency
AGCCGT (Haplotype 1)	0.024
AGTCGT (Haplotype 2)	0.027
GGCCGT (Haplotype 3)	0.195
AATCGT (Haplotype 4)	0.027
AGACGG (Haplotype 5)	0.079
AGCCAT (Haplotype 6)	0.256
AGTTGT (Haplotype 7)	0.241
GGCTGT (Haplotype 8)	0.036
AGATGG (Haplotype 9)	0.036
AGCTAT (Haplotype 10)	0.026

-717A/G is at position 1, -305G/A is at position 2, -286C/T/A is at position 3, +1444C/T is at position 4, +2302G/A is at position 5 and +4899T/G is at position 6.

8.4.2 Baseline characteristics by CRP genotype and haplotype

Baseline clinical characteristics of the subjects divided according to CRP genotype are reported in Tables 8.4-8.9. The number of individuals with -305GA and AA genotypes was very small and for this reason, these genotypes were combined and compared with the GG genotype. For the –286C/T/A polymorphism, the TT, TA, CA and AA genotypes were also combined since previous work had suggested that the T- and A-alleles had the highest CRP concentrations and the number of individuals with A-alleles was small so subjects with T- and A-alleles were grouped together and compared to CC and CT individuals.

For the –717A/G polymorphism, no significant differences were detected between genotypic classes for any traditional cardiovascular risk factors, except systolic and diastolic blood pressure ($p=0.02$, $p=0.008$ respectively) or dental parameters including severity and extent of periodontal infections. When the –305G/A polymorphism was examined, there were significant differences between genotypic groups for proportion of males ($p<0.001$) and ethnicity ($p<0.001$). However, all these differences are most likely due to the small sample sizes for the GA and AA genotypic classes since this site is not polymorphic in all populations. For the +1444C/T polymorphism, no significant differences were seen for any cardiovascular risk factors, except age ($p=0.002$). For the –286C/T/A and +2302G/A polymorphisms, no significant differences were seen between the genotypic classes. When the +4899T/G polymorphism was evaluated, the only small significant difference seen between the TT, TG and GG genotypes was for ethnicity ($p=0.03$). The differences in cardiovascular risk factors seen here by genotype are most likely due to be the consequence of multiple testing resulting in false positive outcomes, since several characteristics were measured, and are not observed in larger studies (see Chapters 6 and 7).

Table 8.4. Baseline characteristics of subjects included in the Periodontal disease study, according to the CRP -717A/G genotype.

Variable	AA (N=93) Mean (SD)	AG (N=56) Mean (SD)	GG (N=13) Mean (SD)	p-value
Age (years)	47.5 (8.1)	45.9 (8.1)	48.0 (5.9)	0.44
Gender, Males, n (%)*	49 (53%)	26 (46%)	6 (46%)	0.73
Ethnicity, Caucasians, n (%)*	57 (61%)	44 (79%)	7 (54%)	0.06
BMI (Kg/m^2)	27.1 (4.0)	26.2 (3.8)	28.5 (5.8)	0.22
Smoking, Current, n (%)*	27 (29%)	18 (32%)	4 (31%)	0.92
Systolic BP (mmHg)	133.8 (17.3)	126.7 (14.0)	124.8 (18.1)	0.02
Diastolic BP (mmHg)	86.2 (11.2)	81.1 (10.0)	80.3 (9.2)	0.008
Number of affected teeth	27.4 (2.9)	27.0 (3.2)	28.0 (2.1)	0.50
Periodontal disease severity†	79.6 (27.9)	74.8 (25.2)	87.8 (29.8)	0.26
IL-6, baseline (pg/mL)	2.1 (4.0)	1.6 (2.6)	2.2 (2.0)	0.70
IL-6, peak (pg/mL)	6.8 (7.0)	6.1 (5.0)	8.2 (6.6)	0.55
CRP, baseline (mg/L)	1.70 (1.41)	1.93 (1.59)	2.21 (2.15)	0.46

A t-test was performed for continuous variables, and a Chi-Square Test (*) was used for dichotomous variables.

†Total number of periodontal lesions.

BMI indicates body mass index, BP indicates blood pressure.

Table 8.5. Baseline characteristics of subjects included in the Periodontal disease study, according to the CRP –305G/A genotype.

Variable	GG (N=149) Mean (SD)	GA+AA (N=12) Mean (SD)	p-value
Age (years)	47.2 (7.9)	43.8 (9.0)	0.16
Gender, Males, n (%)*	72 (48%)	7 (58%)	<0.001
Ethnicity, Caucasians, n (%)*	104 (70%)	2 (17%)	<0.001
BMI (Kg/m^2)	26.7 (4.5)	28.9 (4.3)	0.11
Smoking, Current, n (%)*	46 (31%)	2 (17%)	0.51
Systolic BP (mmHg)	130.6 (15.6)	133.5 (27.2)	0.56
Diastolic BP (mmHg)	83.9 (10.5)	85.8 (15.8)	0.57
Number of affected teeth	27.4 (2.9)	26.8 (3.0)	0.50
Periodontal disease severity†	77.2 (26.8)	88.9 (30.9)	0.15
IL-6, baseline (pg/mL)	1.8 (2.4)	0.7 (1.2)	0.54
IL-6, peak (pg/mL)	6.7 (6.5)	5.6 (5.3)	0.56
CRP, baseline (mg/L)	1.84 (1.58)	1.55 (0.98)	0.74

A t-test was performed for continuous variables, and a Chi-Square Test (*) was used for dichotomous variables.

†Total number of periodontal lesions.

BMI indicates body mass index, BP indicates blood pressure.

Table 8.6. Baseline characteristics of subjects included in the Periodontal disease study, according to the CRP –286C/T/A genotype.

Variable	CC (N=52) Mean (SD)	CT (N=63) Mean (SD)	TT+TA+CA+ AA (N=49) Mean (SD)	p-value
Age (years)	46.8 (7.8)	46.4 (7.6)	47.5 (8.5)	0.76
Gender, Males, n (%)*	27 (52%)	30 (48%)	24 (49%)	0.90
Ethnicity, Caucasians, n (%)*	38 (73%)	43 (68%)	27 (55%)	0.14
BMI (Kg/m^2)	27.2 (4.7)	27.2 (4.1)	26.3 (5.0)	0.50
Smoking, Current, n (%)*	16 (31%)	18 (29%)	15 (31%)	0.96
Systolic BP (mmHg)	127.1 (15.0)	131.0 (15.3)	133.9 (19.1)	0.11
Diastolic BP (mmHg)	80.9 (9.5)	85.3 (9.9)	85.5 (12.7)	0.05
Number of affected teeth	27.4 (3.0)	27.4 (2.7)	27.1 (3.0)	0.86
Periodontal disease severity†	79.2 (27.7)	80.3 (26.4)	75.1 (28.2)	0.59
IL-6, baseline (pg/mL)	1.7 (2.1)	1.8 (2.1)	2.5 (5.4)	0.42
IL-6, peak (pg/mL)	6.6 (4.9)	5.9 (6.2)	7.5 (7.7)	0.40
CRP, baseline (mg/L)	1.78 (1.42)	1.77 (1.54)	1.85 (1.58)	0.96

A t-test was performed for continuous variables, and a Chi-Square Test (*) was used for dichotomous variables.

†Total number of periodontal lesions.

BMI indicates body mass index, BP indicates blood pressure.

Table 8.7. Baseline characteristics of subjects included in the Periodontal disease study, according to the CRP +1444C/T genotype.

Variable	CC (N=71) Mean (SD)	CT (N=66) Mean (SD)	TT (N=25) Mean (SD)	p-value
Age (years)	45.7 (8.1)	46.6 (6.4)	51.9 (9.2)	0.002
Gender, Males, n (%)*	37 (52%)	33 (50%)	11 (44%)	0.78
Ethnicity, Caucasians, n (%)*	45 (63%)	47 (71%)	15 (60%)	0.49
BMI (Kg/m^2)	27.2 (4.9)	27.1 (4.4)	25.9 (3.9)	0.45
Smoking, Current, n (%)*	20 (28%)	23 (35%)	6 (24%)	0.53
Systolic BP (mmHg)	130.3 (16.4)	128.9 (16.0)	135.5 (17.7)	0.23
Diastolic BP (mmHg)	83.4 (11.0)	83.7 (10.6)	85.5 (10.8)	0.70
Number of affected teeth	27.5 (2.8)	27.3 (2.6)	26.8 (3.8)	0.65
Periodontal disease severity†	79.3 (28.3)	79.7 (27.4)	72.9 (24.5)	0.54
IL-6, baseline (pg/mL)	1.7 (2.0)	2.5 (5.0)	1.3 (0.8)	0.27
IL-6, peak (pg/mL)	5.9 (4.8)	7.3 (8.1)	7.1 (5.1)	0.45
CRP, baseline (mg/L)	1.87 (1.58)	1.90 (1.56)	1.51 (1.27)	0.46

A t-test was performed for continuous variables, and a Chi-Square Test (*) was used for dichotomous variables.

†Total number of periodontal lesions.

BMI indicates body mass index, BP indicates blood pressure.

Table 8.8. Baseline characteristics of subjects included in the Periodontal disease study, according to the CRP +2302G/A genotype.

Variable	GG (N=79) Mean (SD)	GA (N=74) Mean (SD)	AA (N=10) Mean (SD)	p-value
Age (years)	47.5 (7.9)	46.0 (8.2)	49.2 (6.6)	0.35
Gender, Males, n (%)*	37 (47%)	39 (53%)	5 (50%)	0.77
Ethnicity, Caucasians, n (%)*	48 (61%)	51 (69%)	9 (90%)	0.15
BMI (Kg/m^2)	26.6 (4.9)	27.4 (4.2)	25.9 (4.7)	0.45
Smoking, Current, n (%)*	26 (33%)	20 (27%)	3 (30%)	0.73
Systolic BP (mmHg)	132.4 (17.4)	127.7 (15.5)	138.9 (14.2)	0.06
Diastolic BP (mmHg)	84.8 (10.8)	82.7 (11.2)	87.4 (8.4)	0.30
Number of affected teeth	27.2 (3.0)	27.4 (2.9)	27.6 (3.0)	0.82
Periodontal disease severity†	79.3 (28.1)	76.6 (25.7)	83.2 (34.2)	0.70
IL-6, baseline (pg/mL)	2.1 (4.3)	1.9 (2.5)	1.7 (1.2)	0.89
IL-6, peak (pg/mL)	6.9 (7.0)	6.5 (6.0)	5.9 (2.5)	0.87
CRP, baseline (mg/L)	2.00 (1.67)	1.64 (1.42)	1.70 (0.83)	0.32

A t-test was performed for continuous variables, and a Chi-Square Test (*) was used for dichotomous variables.

†Total number of periodontal lesions.

BMI indicates body mass index, BP indicates blood pressure.

Table 8.9. Baseline characteristics of subjects included in the Periodontal disease study, according to the CRP +4899T/G genotype.

Variable	TT (N=123) Mean (SD)	TG (N=31) Mean (SD)	GG (N=6) Mean (SD)	p-value
Age (years)	47.4 (8.2)	45.7 (6.8)	46.8 (9.8)	0.59
Gender, Males, n (%)*	63 (51%)	15 (48%)	3 (50%)	0.96
Ethnicity, Caucasians, n (%)*	87 (71%)	16 (52%)	2 (33%)	0.03
BMI (Kg/m^2)	26.9 (4.3)	26.8 (5.3)	27.2 (6.6)	0.98
Smoking, Current, n (%)*	38 (31%)	6 (19%)	4 (67%)	0.06
Systolic BP (mmHg)	130.0 (15.5)	133.3 (22.0)	126.3 (4.8)	0.51
Diastolic BP (mmHg)	83.2 (9.8)	87.4 (15.1)	80.7 (5.2)	0.13
Number of affected teeth	27.4 (2.8)	26.8 (3.6)	27.7 (2.5)	0.57
Periodontal disease severity†	79.3 (27.3)	74.4 (29.1)	72.3 (18.7)	0.59
IL-6, baseline (pg/mL)	1.7 (2.0)	3.1 (6.0)	1.0 (0.8)	0.10
IL-6, peak (pg/mL)	6.3 (5.5)	8.6 (9.2)	5.5 (4.7)	0.16
CRP, baseline (mg/L)	1.79 (1.50)	1.56 (1.20)	3.06 (2.99)	0.19

A t-test was performed for continuous variables, and a Chi-Square Test (*) was used for dichotomous variables.

†Total number of periodontal lesions.

BMI indicates body mass index, BP indicates blood pressure.

Baseline clinical characteristics were also divided according to haplotype (see Table 8.10). Although the number of individuals in some haplotypic groups was very small, no significant differences were seen for any cardiovascular risk factors across the haplotypes.

Table 8.10. Baseline characteristics of subjects included in the Periodontal disease study, according to CRP haplotypes.

	-717A/G//-305G/A//- 286C/T/A//+1444C/T//+2302G/A//+4899T/G										
	AGCCGT (N=4)	**AGTCGT (N=4)**	**GGCCGT (N=32)**	**AATCGT (N=4)**	**AGACGG (N=13)**	**AGCCAT (N=42)**	**AGTTGT (N=39)**	**GGCTGT (N=6)**	**AGATGG (N=6)**	**AGCTAT (N=4)**	**P value**
Baseline CRP[†] (mg/L)	1.25 (0.24-6.58)	3.00 (1.11-8.10)	2.06 (1.34-3.18)	1.28 (0.30-5.42)	3.27 (1.58-6.74)	1.49 (0.93-2.40)	1.70 (1.12-2.57)	4.86 (0.87-27.0)	0.75 (0.17-3.30)	1.21 (0.10-14.5)	0.09
Age[†] (years)	43.6 (27.2-59.8)	56.4 (45.8-67)	48.0 (43.9-52.2)	38.3 (23.8-52.8)	44.9 (37.0-52.8)	46.6 (43-50.2)	51.2 (47.4-54.8)	48.0 (35.4-60.4)	55.0 (44.2-65.9)	63.0 (28.8-97.2)	0.05
BMI[†] (Kg/m^2)	28.7 (22.5-36.7)	23.4 (18.9-28.9)	24.7 (22.6-26.9)	27.7 (20.8-36.9)	27.1 (24.0-30.6)	25.0 (23.0-27.3)	24.8 (22.8-27.0)	28.1 (23.7-33.3)	18.9 (15.223.7)	26.6 (20.6-34.3)	0.05
Systolic BP[†] (mmHg)	156.5 (128.5-190.6)	145.2 (119.8-175.9)	127.7 (120.5- 135.4)	121.0 (100.5-145.5)	131.6 (119.2-145.5)	136.0 (128.3-144.3)	138.6 (130.1-147.5)	122.5 (103.3-145.3)	123.3 (104.6-145.5)	137.6 (100.9-187.2)	0.39
Smoking[‡]	1.72 (0.24-12.58)	2.69 (0.43-16.65)	1.01 (0.44-2.31)	0.57 (0.06-5.22)	1.81 (0.65-5.00)	1.00	1.36 (0.65-2.88)	3.91 (0.64-2.88)	2.00 (0.41-9.72)	0.57 (0-612)	0.87

[†]Geometric mean.

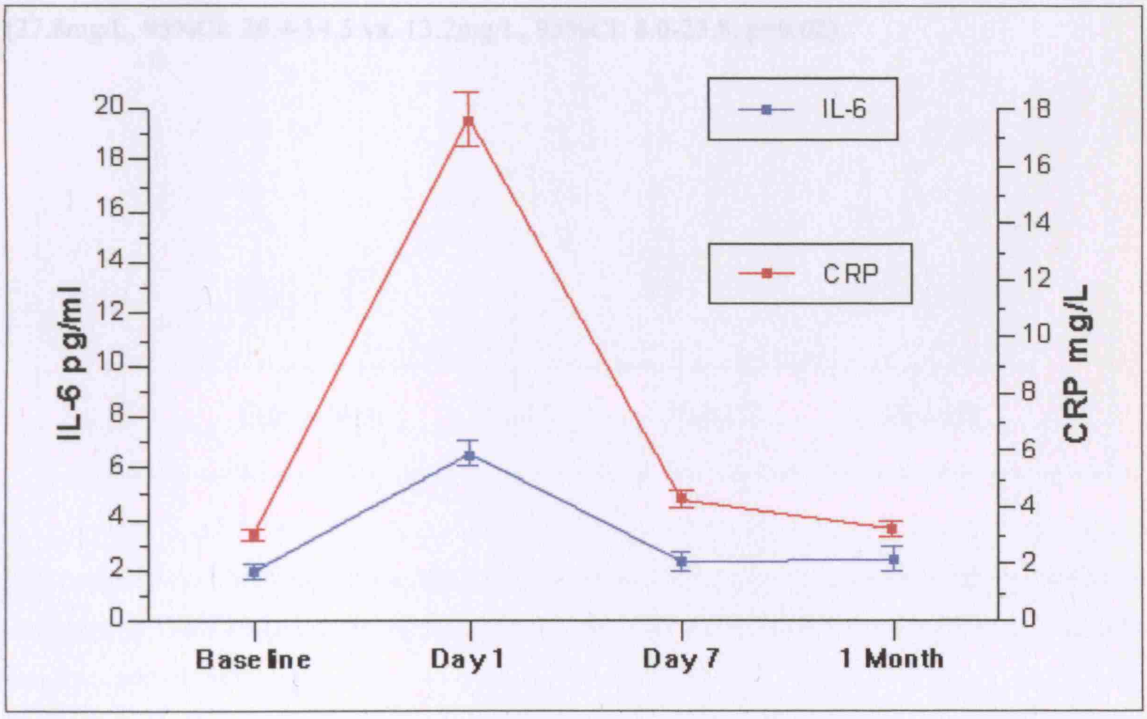
[‡]Odds ratio for smoking using the AGCCAT haplotype as the reference.

All values are for two copies of the haplotype.

8.4.3 Effect of periodontal treatment on the inflammatory response

The inflammatory stimulus (intensive periodontal therapy) resulted in a sharp significant increase of CRP on day one, an increase of 11.7mg/L compared to baseline concentrations ($p<0.0001$). CRP concentrations then started to fall and by day seven, the CRP concentration had fallen, but was still significantly higher than baseline (mean difference of 1.4mg/L, $p<0.0001$) (Figure 8.2). The difference in CRP concentration between baseline and day one and day seven remained statistically significant in the multivariate analysis ($p<0.0001$) after adjusting for age, gender, ethnicity, body mass index, smoking, blood pressure, log [IL-6] and number of teeth extracted during treatment; (corrected model $R^2=0.68$). IL-6 concentrations also significantly increased on day one compared to baseline, then started to decrease and had returned to usual concentrations by day seven.

Figure 8.2. Changes in CRP and IL-6 before and after periodontal therapy.



IL-6 and CRP mean values and standard errors at each visit.
Values are based on 164 subjects with both IL-6 and CRP measures.

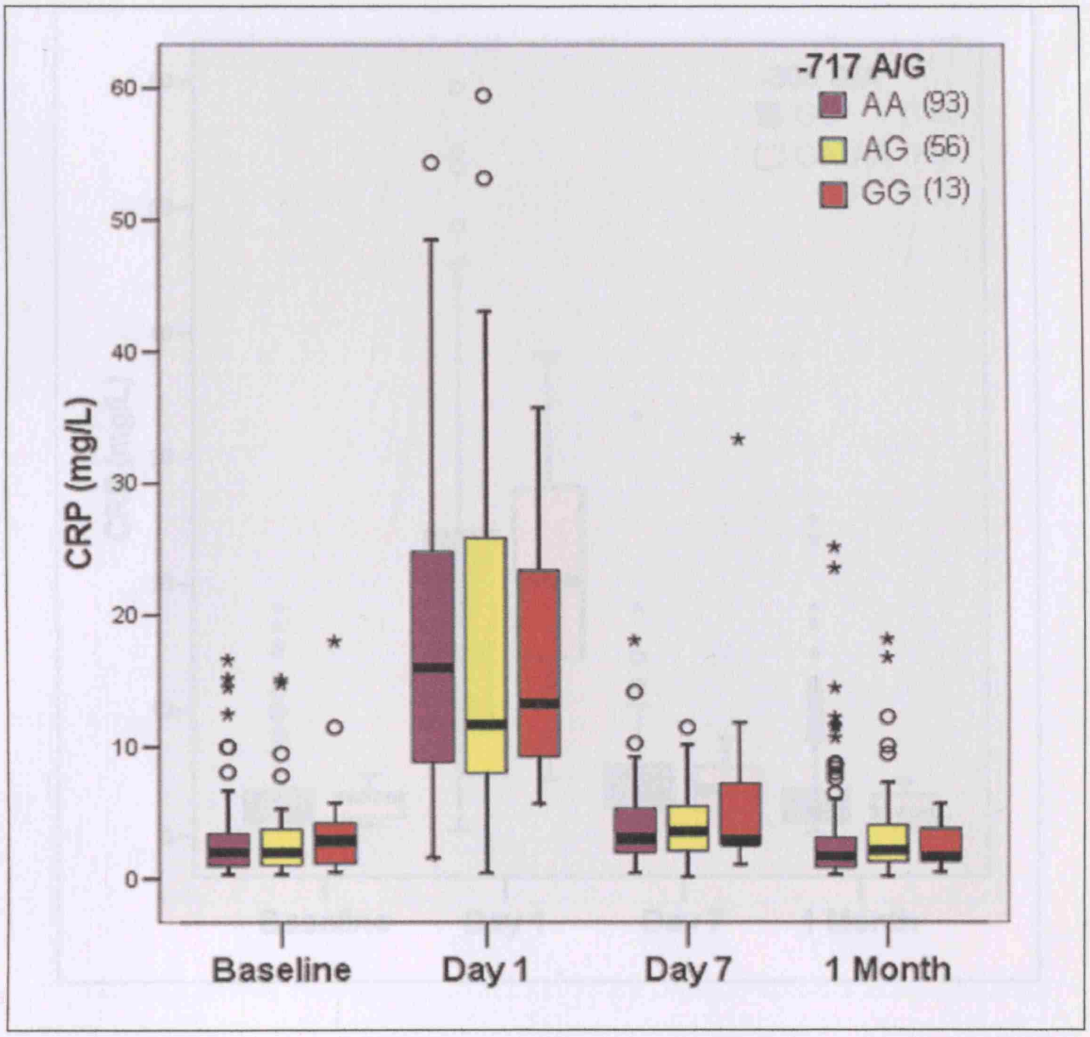
8.4.4 Effect of genotype and haplotype on the CRP response to periodontal treatment

With the exception of the –717A/G polymorphism, the magnitude of the acute elevation of CRP was dependent on CRP genotype (Figures 8.3-8.9). For the -305G/A polymorphism, A-allele carriers had higher concentrations of CRP compared to subjects homozygous for the G-allele at day one (20.1mg/L, 95%CI: 13.6-27.5 vs. 14.0mg/L, 95%CI: 8.5-24.4, $p=0.61$). When

the –286C/T/A polymorphism was examined, the TT, TA, CA and AA genotypes were combined and compared with the CC genotypes (see Figure 8.6) and showed a significant difference at day one (20.75mg/L, 95%CI: 13.0-36.3 vs. 11.0mg/L, 95%CI: 8.4-21.4, p=0.03). CRP values according to individual –286C/T/A genotypes also differed and are shown in Figure 8.5.

For the +1444C/T polymorphism, subjects homozygous for the T-allele had a significantly higher concentration of CRP compared to C-allele homozygotes at day one (20.5mg/L, 95%CI: 10.5-31.0 vs. 17.0mg/L, 95%CI: 9.8-24.3, p=0.01) and differences persisted but were smaller by day seven. For the +2302G/A polymorphism, subjects homozygous for the rare A-allele had a lower concentration of CRP compared to subjects homozygous for the G-allele at day one (9.4mg/L, 95%CI: 8.5-17.0 vs. 18mg/L, 95%CI: 9.5-27.5, p=0.02). When the +4899T/G polymorphism was examined, subjects homozygous for the rare G-allele also had higher concentrations of CRP compared to subjects homozygous for the T-allele at day one (27.8mg/L, 95%CI: 26.4-34.5 vs. 13.2mg/L, 95%CI: 8.0-23.8, p=0.02).

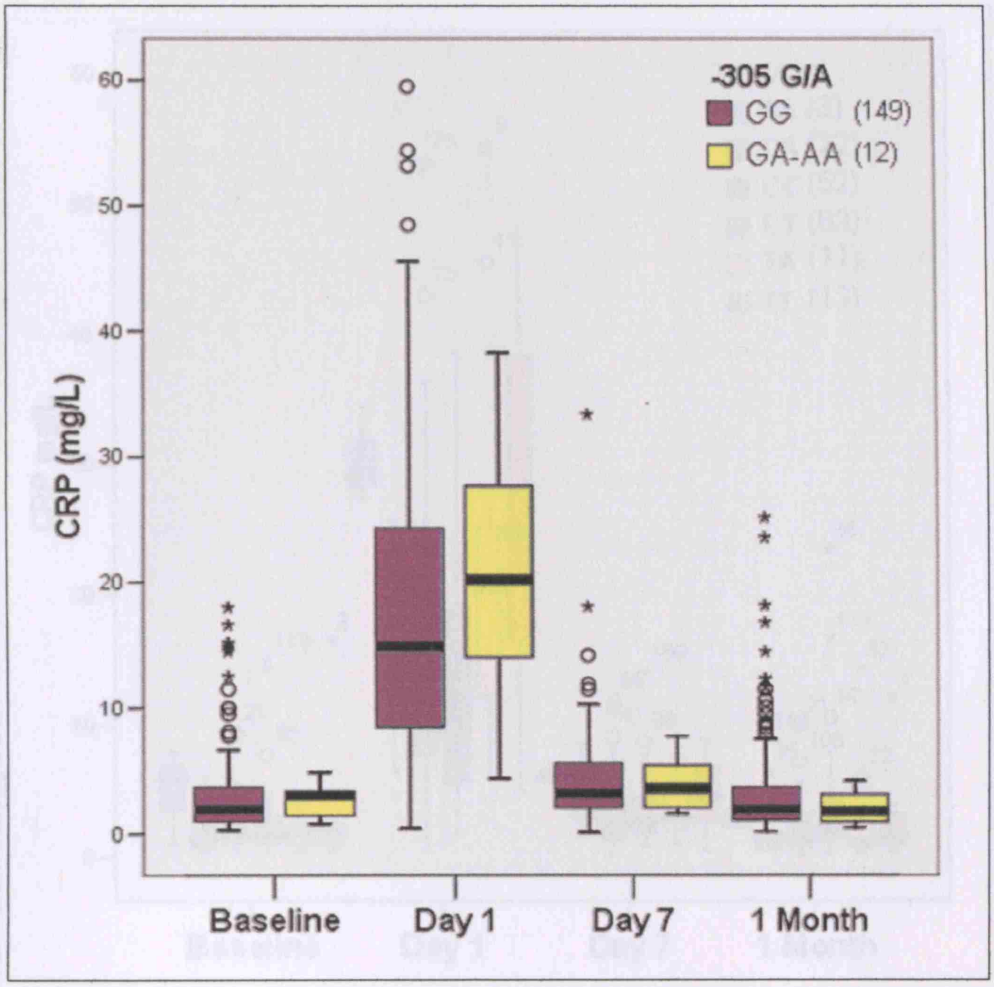
Figure 8.3. Box and whisker plots showing changes in CRP before and after periodontal therapy according to –717A/G genotypes.



The boundaries of the box refer to the 25th (lower) and 75th (upper) percentiles and the median is indicated by the horizontal line inside the box. Open circles represent outliers whereas asterisks stand for extreme observations.

Difference in CRP concentrations by genotype at day 1: 13.5 (9.5-23.5) vs. 16.0 (9.0-24.8) for GG vs. AA genotypes, p=0.97 by ANOVA.

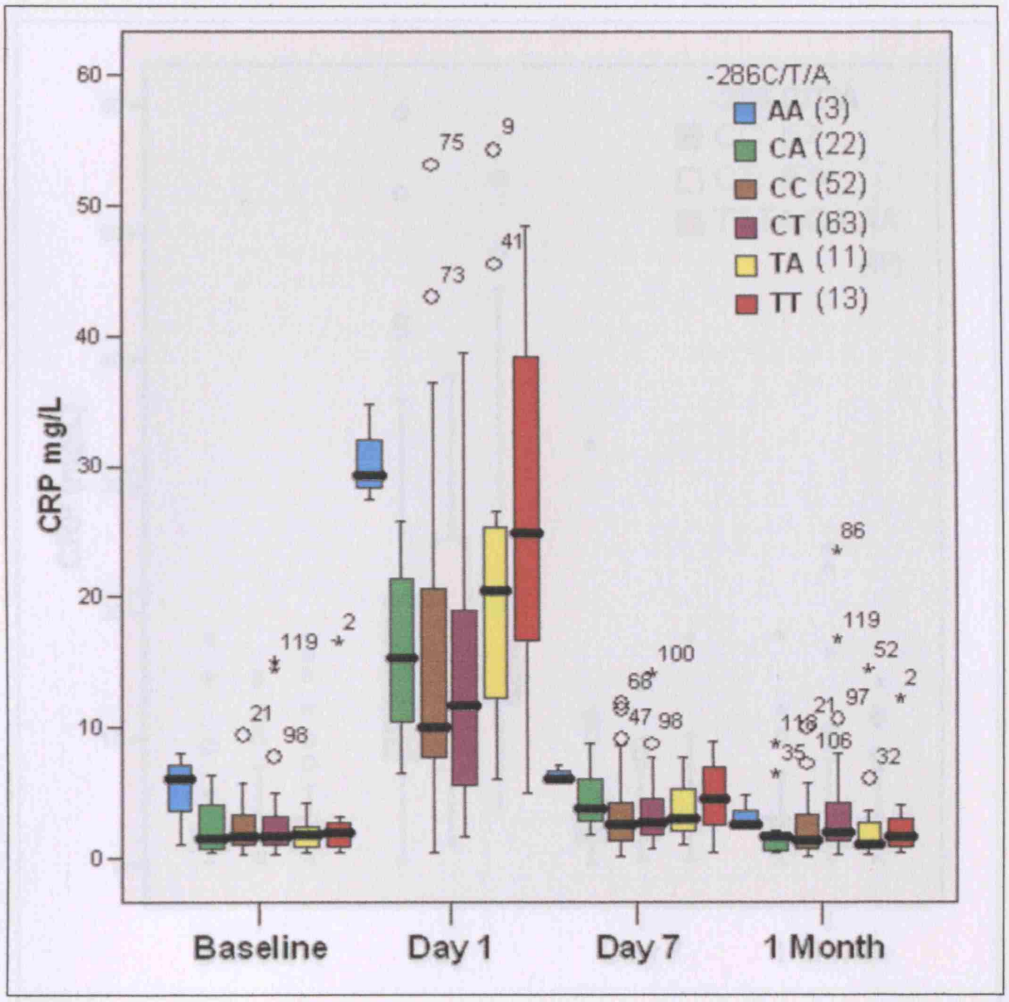
Figure 8.4. Box and whisker plots showing changes in CRP before and after periodontal therapy according to –305GG and A-allele carriers.



The boundaries of the box refer to the 25th (lower) and 75th (upper) percentiles and the median is indicated by the horizontal line inside the box. Open circles represent outliers whereas asterisks stand for extreme observations.

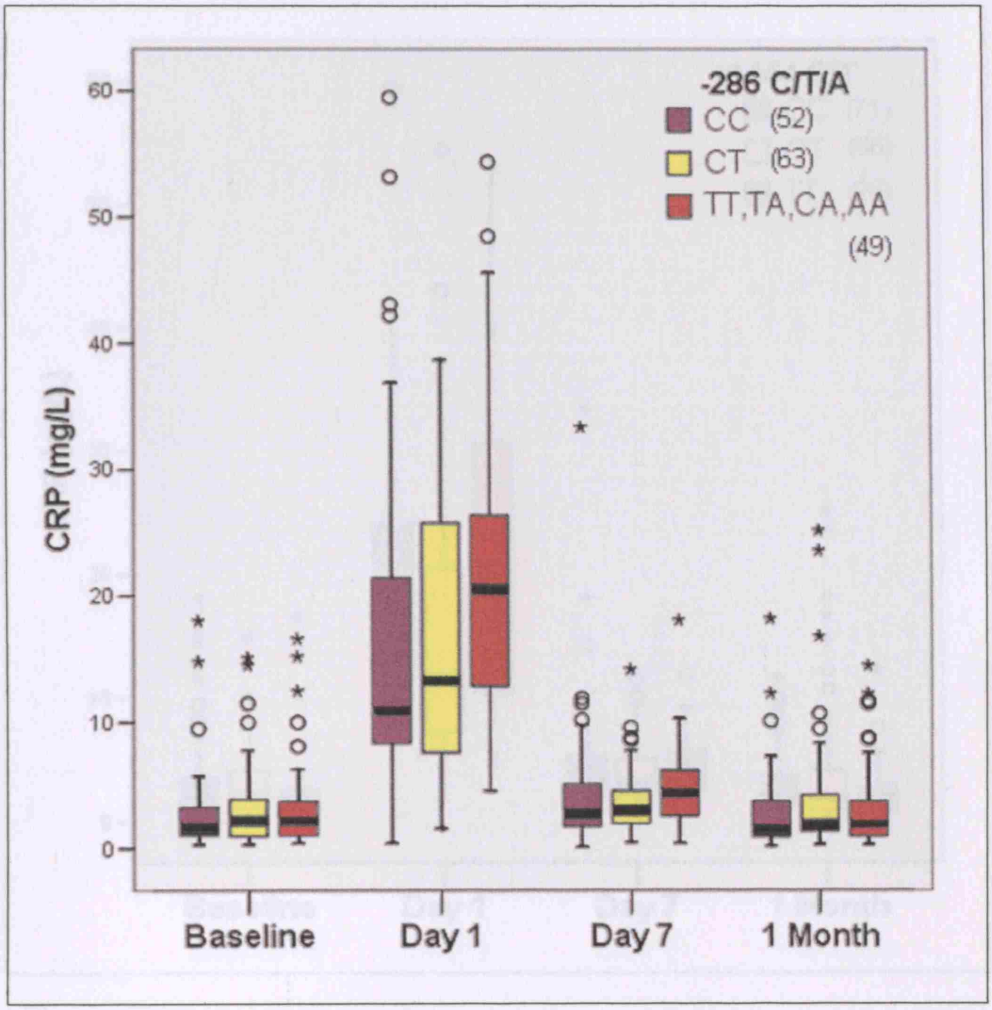
Difference in CRP concentrations by genotype at day 1: 20.1 (13.6-27.5) vs. 14.0 (8.5-24.4) for A-allele carriers vs. GG genotypes, p=0.61 by ANOVA.

Figure 8.5. Box and whisker plots showing changes in CRP before and after periodontal therapy according to –286C/T/A genotypes.



The boundaries of the box refer to the 25th (lower) and 75th (upper) percentiles and the median is indicated by the horizontal line inside the box. Open circles represent outliers whereas asterisks stand for extreme observations.

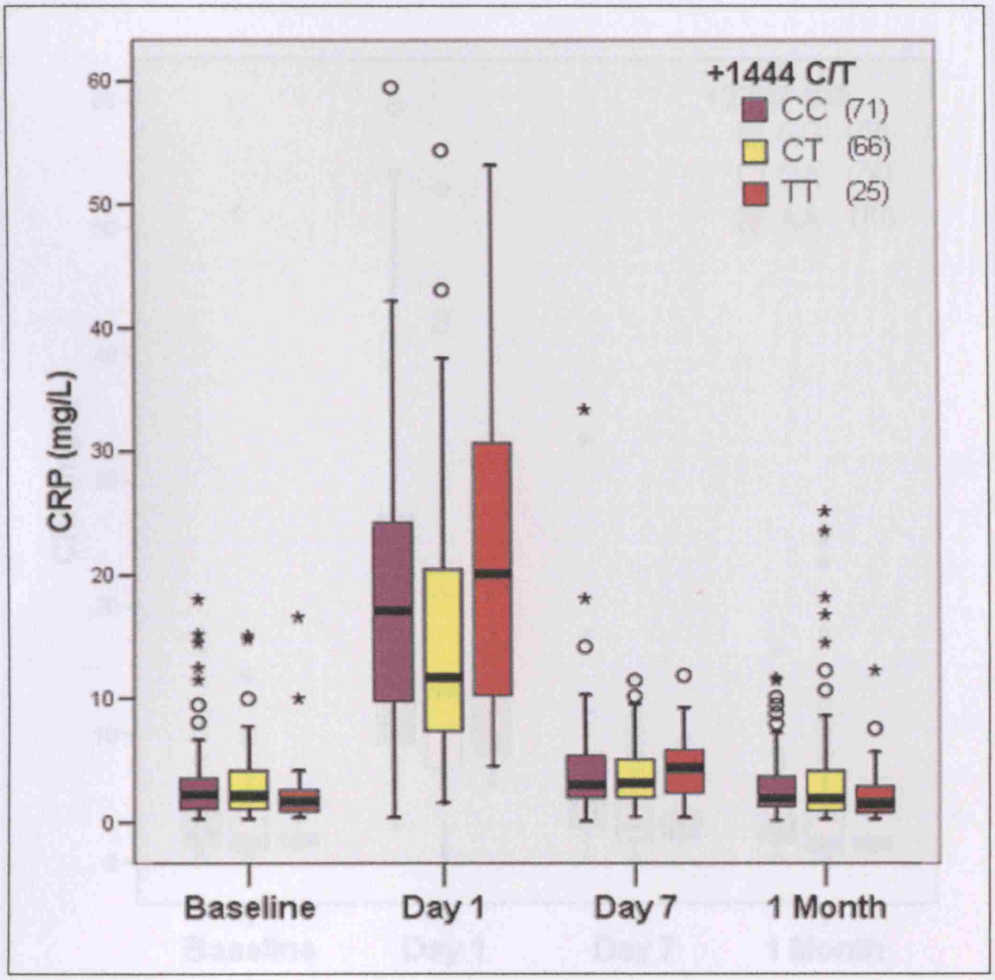
Figure 8.6. Box and whisker plots showing changes in CRP before and after periodontal therapy according to –286CC, CT and (TT, TA, CA, AA) genotypes.



The boundaries of the box refer to the 25th (lower) and 75th (upper) percentiles and the median is indicated by the horizontal line inside the box. Open circles represent outliers whereas asterisks stand for extreme observations.

Difference in CRP concentrations by genotype at day 1: 20.75 (13.0-36.3) vs. 11.0 (8.4-21.4) for (TT, TA, CA, AA) vs. CC genotypes, p=0.03 by ANOVA.

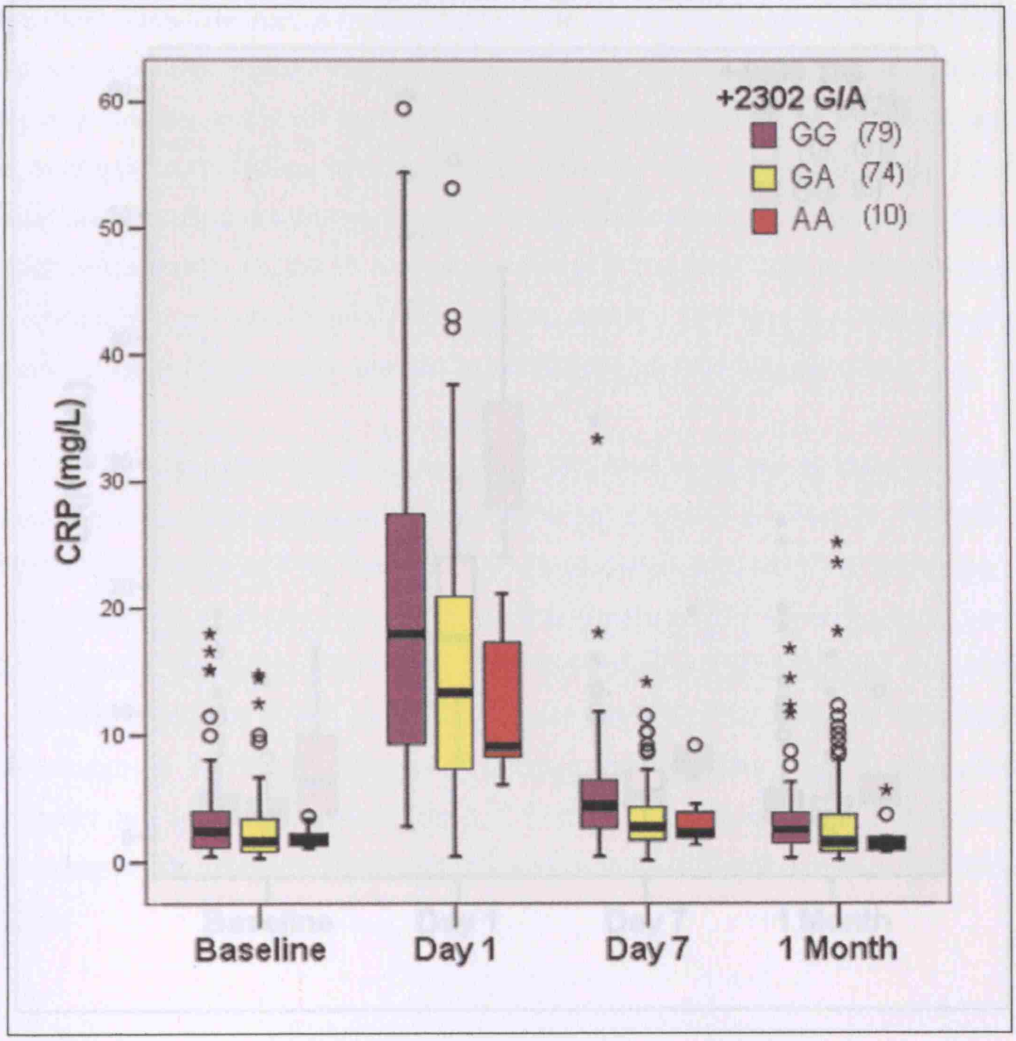
Figure 8.7. Box and whisker plots showing changes in CRP before and after periodontal therapy according to +1444C/T genotypes.



The boundaries of the box refer to the 25th (lower) and 75th (upper) percentiles and the median is indicated by the horizontal line inside the box. Open circles represent outliers whereas asterisks stand for extreme observations.

Difference in CRP concentrations by genotype at day 1: 20.5 (10.5-31.0) vs. 17.0 (9.8-24.3) for TT vs. CC genotypes, p=0.01 by ANOVA.

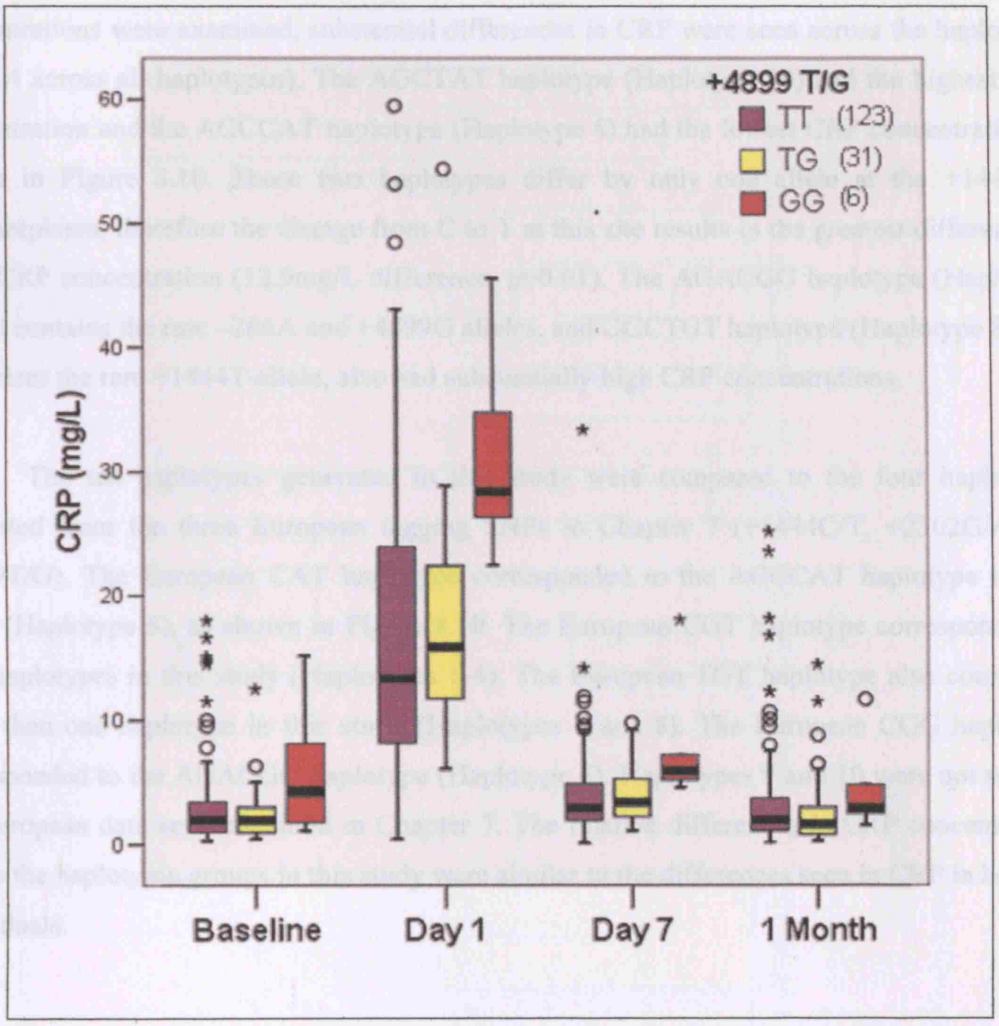
Figure 8.8. Box and whisker plots showing changes in CRP before and after periodontal therapy according to +2302G/A genotypes.



The boundaries of the box refer to the 25th (lower) and 75th (upper) percentiles and the median is indicated by the horizontal line inside the box. Open circles represent outliers whereas asterisks stand for extreme observations.

Difference in CRP concentrations by genotype at day 1: 9.4 (8.5-17.0) vs. 18.0 (9.5-27.5) for AA vs. GG genotypes, p=0.02 by ANOVA.

Figure 8.9. Box and whisker plots showing changes in CRP before and after periodontal therapy according to +4899T/G genotypes.



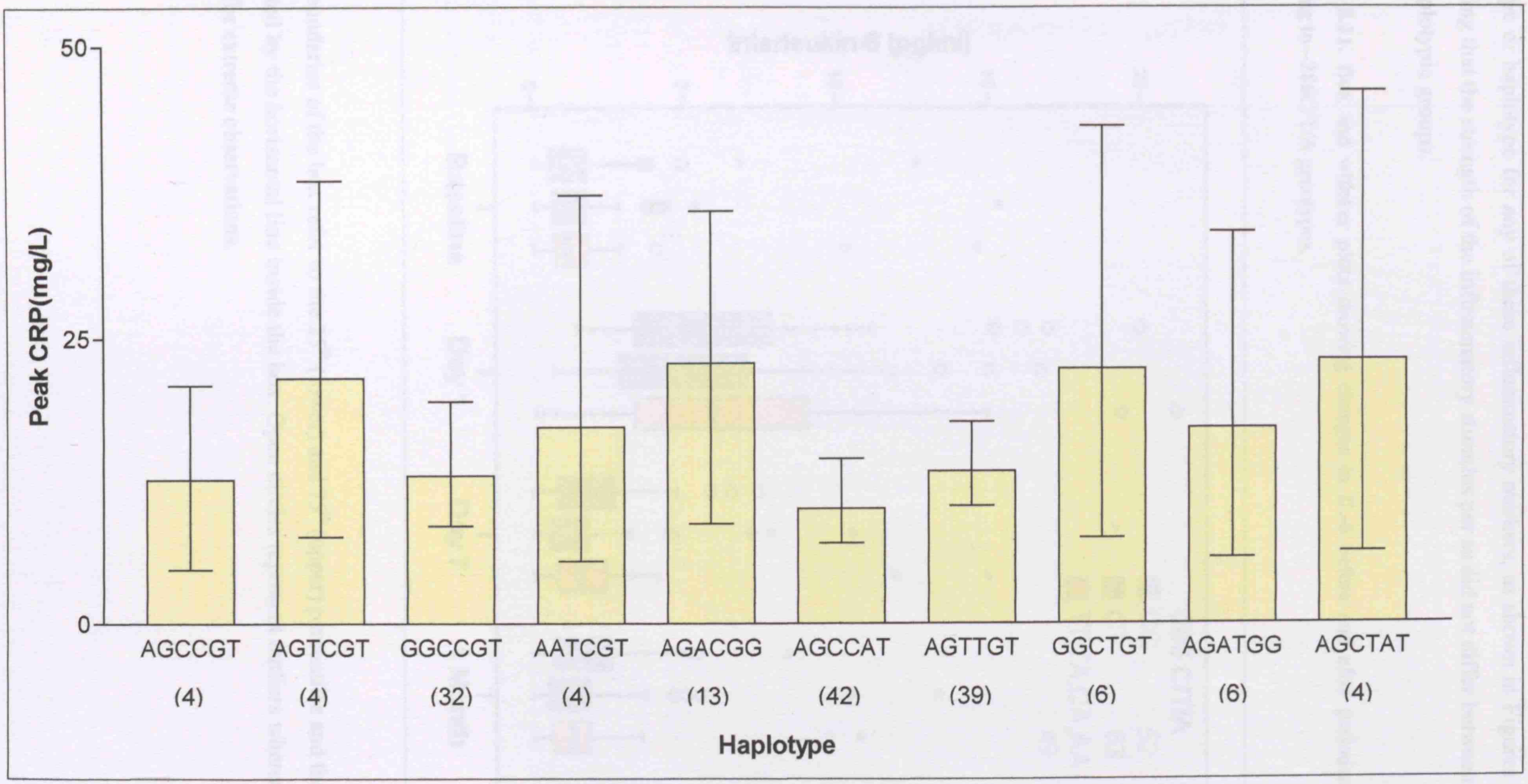
The boundaries of the box refer to the 25th (lower) and 75th (upper) percentiles and the median is indicated by the horizontal line inside the box. Open circles represent outliers whereas asterisks stand for extreme observations.

Difference in CRP concentrations by genotype at day 1: 27.8 (26.4-34.5) vs. 13.2 (8.0-23.8) for GG vs. TT genotypes, p=0.02 by ANOVA.

The six tagging SNPs were then used to generate ten haplotypes with a frequency greater than 0.02 and the effect of haplotypes on CRP concentration was assessed. Although no significant differences were seen for baseline CRP concentrations, when peak CRP concentrations were examined, substantial differences in CRP were seen across the haplotypes (p=0.04 across all haplotypes). The AGCTAT haplotype (Haplotype 10) had the highest CRP concentration and the AGCCAT haplotype (Haplotype 6) had the lowest CRP concentration as shown in Figure 8.10. These two haplotypes differ by only one allele at the +1444C/T polymorphism, therefore the change from C to T at this site results in the greatest difference in peak CRP concentration (12.9mg/L difference, p=0.01). The AGACGG haplotype (Haplotype 5) that contains the rare –286A and +4899G alleles, and GGCTGT haplotype (Haplotype 8) that comprises the rare +1444T-allele, also had substantially high CRP concentrations.

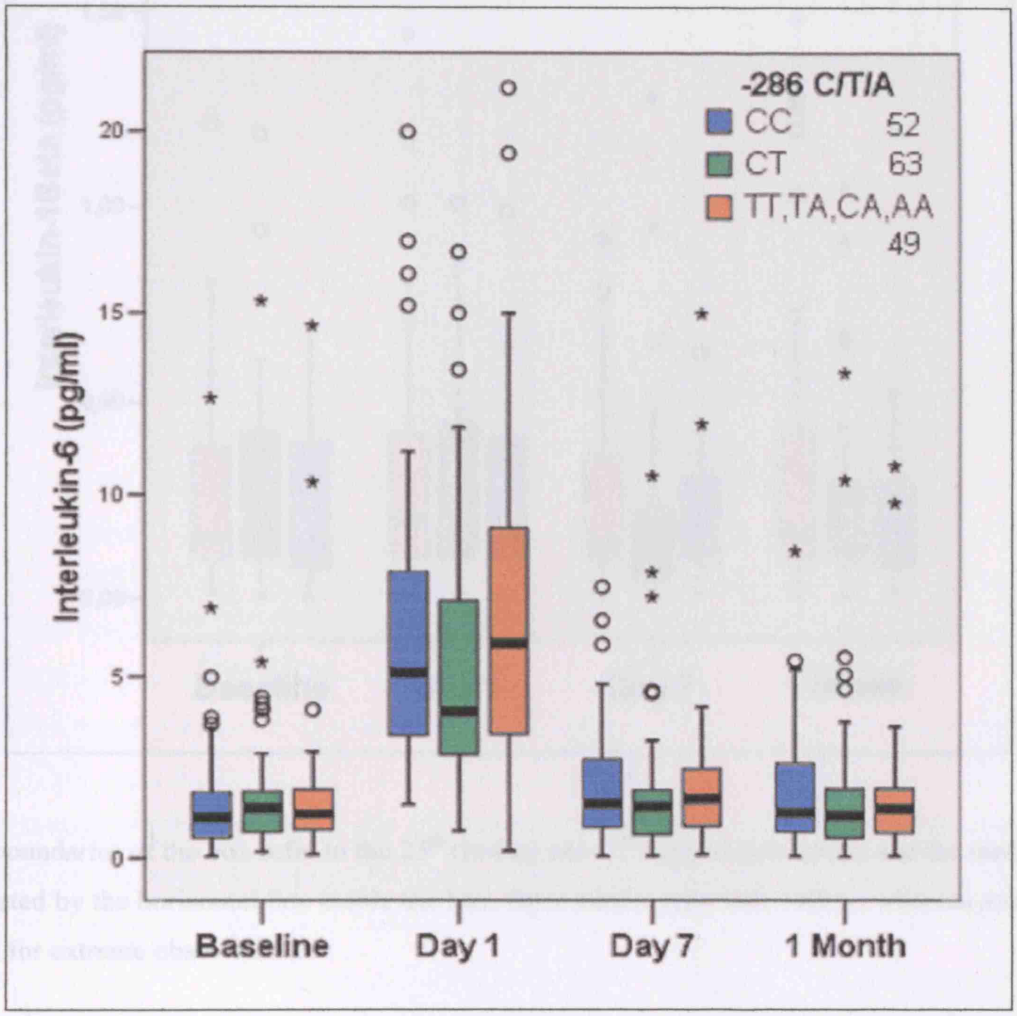
The ten haplotypes generated in this study were compared to the four haplotypes generated from the three European tagging SNPs in Chapter 7 (+1444C/T, +2302G/A and +4899T/G). The European CAT haplotype corresponded to the AGCCAT haplotype in this study (Haplotype 6), as shown in Figure 8.10. The European CGT haplotype corresponded to four haplotypes in this study (Haplotypes 1-4). The European TGT haplotype also comprised more than one haplotype in this study (Haplotypes 7 and 8). The European CGG haplotype corresponded to the AGACGG haplotype (Haplotype 5). Haplotypes 9 and 10 were not seen in the European data sets examined in Chapter 7. The relative differences in CRP concentration across the haplotypic groups in this study were similar to the differences seen in CRP in healthy individuals.

Figure 8.10. Peak CRP concentration (95% CI) according to CRP haplotypes.



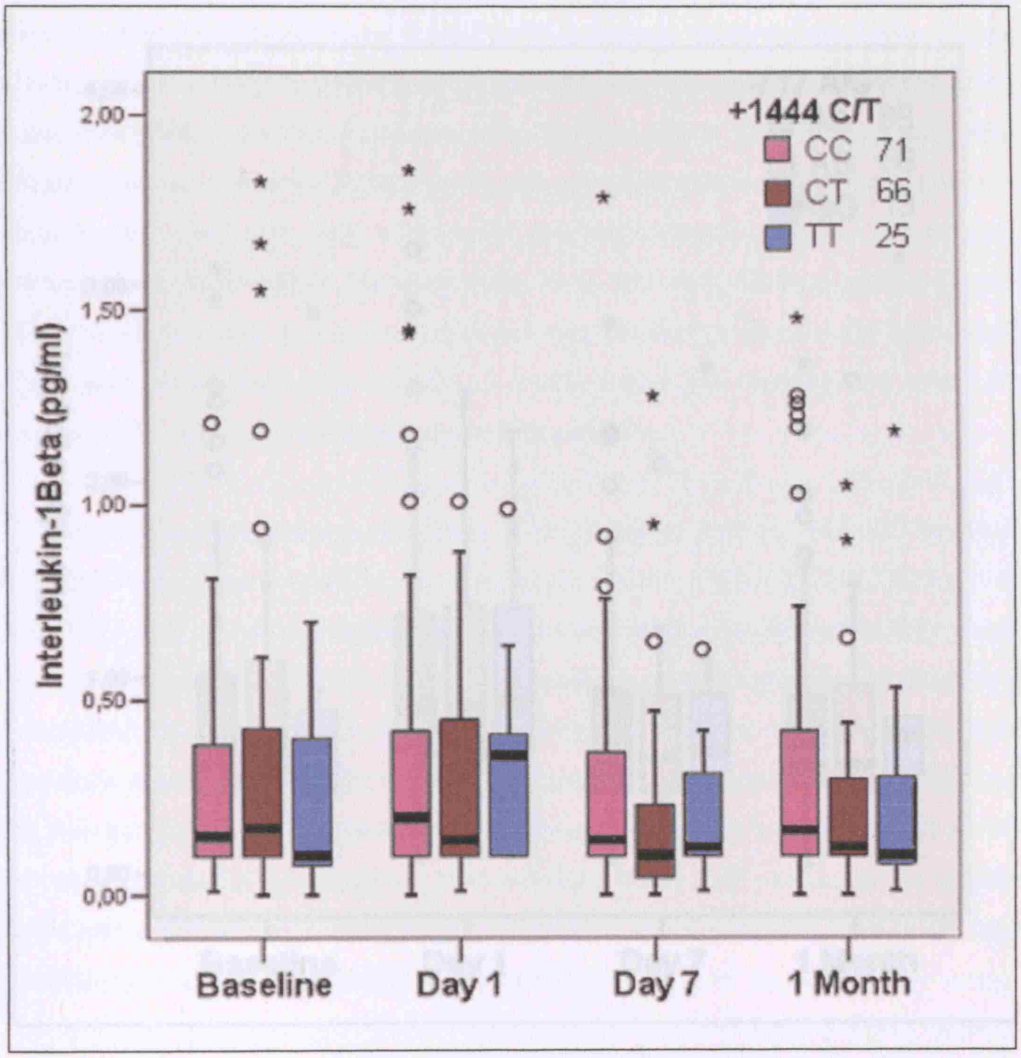
When the effects of genotype and haplotype on the IL-6, IL-1β and TNF-α responses to periodontal treatment were examined, no significant differences were seen according to genotype or haplotype for any of these inflammatory markers, as shown in Figures 8.11-8.13, indicating that the strength of the inflammatory stimulus per se did not differ between genotypic and haplotypic groups.

Figure 8.11. Box and whisker plots showing changes in IL-6 before and after periodontal therapy according to –286C/T/A genotypes.



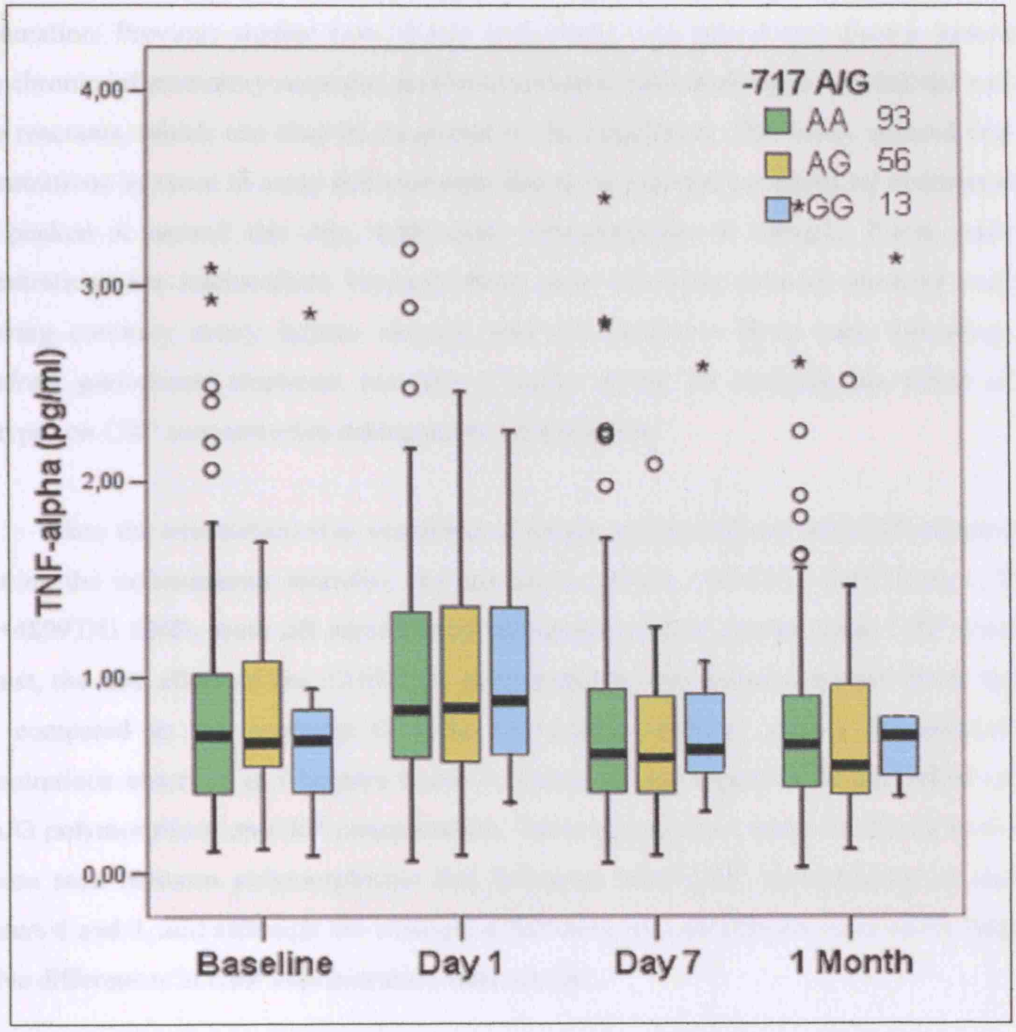
The boundaries of the box refer to the 25th (lower) and 75th (upper) percentiles and the median is indicated by the horizontal line inside the box. Open circles represent outliers whereas asterisks stand for extreme observations.

Figure 8.12. Box and whisker plots showing changes in IL-1β before and after periodontal therapy according to +1444C/T genotypes.



The boundaries of the box refer to the 25th (lower) and 75th (upper) percentiles and the median is indicated by the horizontal line inside the box. Open circles represent outliers whereas asterisks stand for extreme observations.

Figure 8.13. Box and whisker plots showing changes in TNF-α before and after periodontal therapy according to –717A/G genotypes.



The boundaries of the box refer to the 25th (lower) and 75th (upper) percentiles and the median is indicated by the horizontal line inside the box. Open circles represent outliers whereas asterisks stand for extreme observations.

8.5 Discussion

CRP concentration can rise from low basal levels to very high levels in infection and inflammation. Previous studies have shown individuals with periodontal disease have a low-grade chronic inflammatory response, and its treatment results in changes in cytokines and acute phase reactants, which can thus be measured in the circulation. This study showed that CRP concentrations increase in acute inflammation due to periodontal treatment by instrumentation. CRP peaked at around day one, with mean concentrations of 18mg/L. These peak CRP concentrations are intermediate between those seen following military exercise and those following coronary artery bypass surgery, and are similar to those seen following ACS. Therefore, periodontal treatment provides a useful model for studying the effect of CRP genotypes on CRP concentration during acute inflammation.

When the association was examined of single polymorphisms and CRP concentration following the inflammatory stimulus, the rare alleles of the –305G/A, -286C/T/A, +1444C/T and +4899T/G SNPs were all significantly associated with a greater acute CRP release. In contrast, the rare allele of the +2302G/A polymorphism was associated with lower levels of CRP compared to the common G-allele, in keeping with its effects on baseline CRP concentrations observed in Chapters 6 and 7. There did not appear to be any effect of the –717A/G polymorphism on CRP concentration. These associations were consistent in direction to those seen between polymorphisms that influence basal CRP concentration as shown in Chapters 6 and 7, and although the absolute differences in concentration were much larger, the relative differences in CRP concentration were similar.

This suggests that the effect of CRP genotype on CRP concentration at baseline may work through similar mechanisms to those during acute inflammation. Individuals with some genotypes, such as the –286AA, +2302AA and +4899GG variants showed extreme differences in CRP concentration of up to 13.7mg/L compared to other genotypic groups, that might influence diagnostic decisions or prognostic estimates. The associations seen here between single SNPs and peak CRP concentration were similar to the associations seen in a recent study between the –286C/T/A, +1444C/T and +2302G/A CRP polymorphisms and CRP concentration following ACS (Suk Danik *et al.* 2006).

The differences in peak CRP concentrations by genotype were not influenced by systemic differences in conventional cardiovascular risk factors such as age and gender, and inflammatory factors such as IL-6 that are known to affect CRP concentrations. However, peak IL-6 concentrations may have been missed by the sampling time points used, as no measures were taken between baseline and day one, and IL-6 usually peaks a few hours following and

inflammatory stimulus. Therefore, it was not possible to definitely determine that IL-6 did not influence the genotype-CRP association.

The effect of haplotypes generated from the tagging SNPs on peak CRP concentration was also examined. Substantial differences in CRP concentration were seen across the ten haplotypes identified with a frequency greater than 0.02 in this multi-ethnic study. The AGCTAT haplotype (Haplotype 10) exhibited the highest peak CRP concentration and the AGCCAT haplotype (Haplotype 6) the lowest peak CRP concentration. These two haplotypes differ at only one allele (+1444C/T) suggesting that the combination of the remaining alleles together with the change at the +1444C/T site results in the greatest difference in CRP concentration.

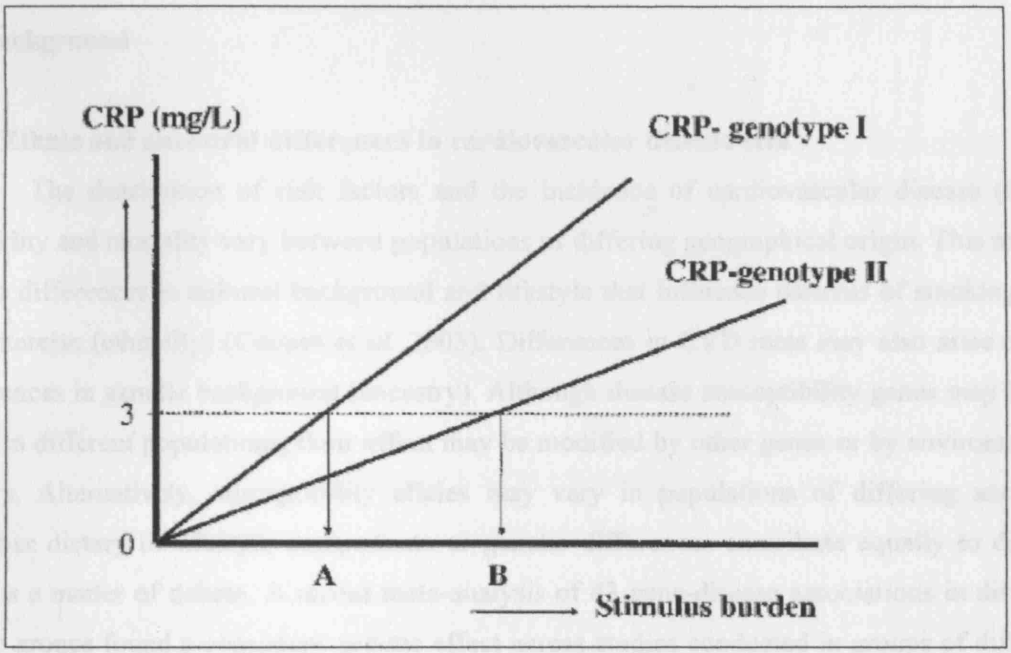
The ten haplotypes generated in this study were compared to the four haplotypes generated from the three European tagging SNPs in Chapter 7 (+1444C/T, +2302G/A and +4899T/G). The European CAT haplotype corresponded to the AGCCAT haplotype in this study (Haplotype 6) and also had the lowest CRP concentrations. The European CGT haplotype corresponded to four haplotypes in this study (Haplotypes 1-4). The European TGT haplotype also comprised more than one haplotype in this study (Haplotypes 7 and 8). The European CGG haplotype corresponded to the AGACGG haplotype (Haplotype 5), which had one of the highest CRP concentrations compared to other haplotypes.

It was surprising that the haplotypes containing the +4899G-allele had a limited effect on CRP concentration since this SNP had a strong effect on CRP concentration in single SNP analyses. Similarly, the haplotypes containing the –286T- and A-alleles also had a smaller effect on CRP concentration to that predicted from single SNP analyses. However, when a large number of haplotypes are identified in a small study, the number of individuals with each haplotype may be small, and the confidence interval limits surrounding the effect may be large, making inference more difficult.

This, and other work showing robust genetic influences on CRP concentration also has relevance for the use of CRP in coronary risk prediction (see Chapter 4). Since CRP concentrations are subject to significant genetic regulation (based on work in this chapter and chapter 6), then the prognostic value of a particular CRP concentration may vary depending on an individual's genotype or haplotype. Two different individuals with the same CRP concentration but contrasting CRP genotype may have very different underlying levels of inflammation. Therefore, the implications of a given CRP value in the diagnosis of an infective or inflammatory disorder, the assessment of treatment efficacy, or the evaluation of prognosis or

outcome could be importantly different in subjects of contrasting genotype (see Figure 8.14). Further studies would be required to evaluate the clinical implication of this finding.

Figure 8.14. Response of two hypothetical individuals with different CRP genotypes who generate different responses of CRP to an inflammatory stimulus (Kluft & de Maat 2003). For any given measured CRP value, the underlying inflammatory or infective burden may differ despite the same measured CRP value.



8.6 Conclusions

The work in this chapter showed that CRP is subject to genetic modulation during acute inflammation. This study has potential mechanistic and clinical relevance. The data suggest that the effects of genotype on CRP concentration are preserved during fluctuations of CRP seen during inflammation. They suggest that common mechanisms may influence CRP concentration in health and in infection or inflammation. The findings also suggest that the use of CRP as a diagnostic test or to monitor treatment may be enhanced by information on genotype. To illustrate, two individuals with the same measured CRP concentration but with contrasting genotype might have substantial differences in the degree of underlying inflammation (as discussed in Chapter 4). Genotype adjusted CRP values may help in the diagnosis of acute inflammation or infective episodes and in the monitoring of their progress.

9. Ancestral differences in CRP

9.1 Aim

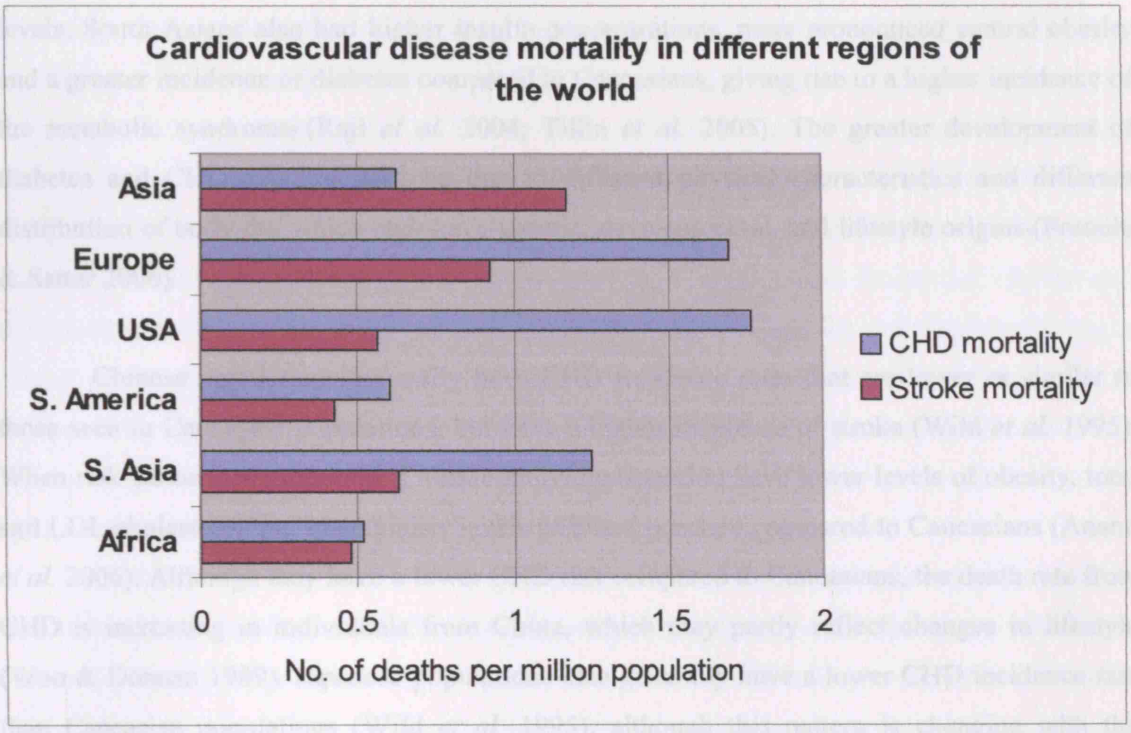
To determine if ethnicity and ancestry influence between population differences in CRP concentration. To assess whether the frequency of CRP genotypes or their effect on CRP concentration differ between groups of differing ancestry.

9.2 Background

9.2.1 Ethnic and ancestral differences in cardiovascular disease risk

The distribution of risk factors and the incidence of cardiovascular disease (CVD) morbidity and mortality vary between populations of differing geographical origin. This may be due to differences in cultural background and lifestyle that influence patterns of smoking, diet and exercise (ethnicity) (Cooper *et al.* 2003). Differences in CVD rates may also arise due to differences in genetic background (ancestry). Although disease susceptibility genes may be the same in different populations, their effect may be modified by other genes or by environmental factors. Alternatively, susceptibility alleles may vary in populations of differing ancestry. Whether dietary or lifestyle components or genetic differences contribute equally to disease rates is a matter of debate. A recent meta-analysis of 43 gene-disease associations in different ethnic groups found a consistent genetic effect across studies conducted in groups of differing geographical region for most associations (Ioannidis *et al.* 2004). Migration followed by the clustering of ethnic minorities among lower socio-economic groups and adoption of a lifestyle of the host region may also contribute to differences in disease risk (Lip & Boos 2005).

Figure 9.1. Cardiovascular disease mortality in population groups of different geographical regions that reflect in part, features of common lifestyles and common ancestry. (Data adapted from the World Health Organisation website, http://www.who.int/cardiovascular_diseases/resources/atlas/ en/ index.html).



Asia refers to China and Japan; Europe refers to France, Sweden and UK; South Asia refers to India and Pakistan.

Although people of African origin living in the UK have a higher incidence of diabetes, hypertension and stroke than those of European background, they appear to have a lower incidence of CHD (Cappuccio *et al.* 1997; Zoratti 1998). In the Atherosclerosis Risk in Communities (ARIC) study, African American people, in contrast, exhibited a different pattern of risk. The CHD mortality rate in men was similar to that seen for American Caucasians, but was higher in African American women (Jones *et al.* 2002). Although African populations tend to have higher blood pressure, BMI and a higher incidence of diabetes, there are also marked differences in the levels of lipoproteins. Africans have higher HDL-cholesterol concentrations and lower triglyceride levels compared to Caucasians, which could provide some protection (Zoratti *et al.* 2000; Howard *et al.* 2003).

Risk factor clustering has been observed in other ancestral groups including South Asians, who have a high incidence of CHD compared to Caucasians, although there appears to be substantial heterogeneity among South Asians from different geographical backgrounds. This

was shown in the Southall study, where conventional risk factors such as hypertension, triglyceride concentrations and smoking rates varied considerably between sub-groups of South Asians (McKeigue *et al.* 1991). When examined together, South Asian individuals with a similar BMI to Caucasian individuals had higher systolic blood pressure and higher triglyceride levels. South Asians also had higher insulin concentrations, more pronounced central obesity and a greater incidence of diabetes compared to Caucasians, giving rise to a higher incidence of the metabolic syndrome (Raji *et al.* 2004; Tillin *et al.* 2005). The greater development of diabetes and CHD may, in part, be due to different physical characteristics and different distribution of body fat, which may have genetic, developmental, and lifestyle origins (Forouhi & Sattar 2006).

Chinese populations generally have CHD incidence rates that are lower or similar to those seen in Caucasian populations, but have a higher incidence of stroke (Wild *et al.* 1995). When risk factors are examined, Chinese individuals tend to have lower levels of obesity, total and LDL-cholesterol, but have higher levels of blood pressure compared to Caucasians (Anand *et al.* 2006). Although they have a lower CHD risk compared to Caucasians, the death rate from CHD is increasing in individuals from China, which may partly reflect changes in lifestyle (Woo & Donnan 1989). Japanese populations also generally have a lower CHD incidence rate than Caucasian populations (Wild *et al.* 1995), although this pattern is changing with the adoption of a more Westernised lifestyle with higher levels of obesity and higher rates of smoking. The result of these lifestyle changes can be seen with higher rates of CHD incidence and morbidity among westernised Japanese living in Hawaii compared to those living in Japan due to higher total and LDL-cholesterol levels, higher BMI and lower HDL-cholesterol levels (Ueshima *et al.* 2003).

Mexican Americans are also part of an ancestral group with higher levels of cardiovascular risk factors including obesity and blood pressure levels, although studies on whether this group has a higher rate of CVD mortality has produced conflicting results. The San Louis Valley Diabetes study reported a lower CVD mortality rate in Mexican Americans and the Corpus Christi study found a higher rate compared to Caucasians (Swenson *et al.* 2002; Goff *et al.* 1997). The San Antonio Heart study, which has been carried out more recently, has compared US-born Mexican Americans, Mexico-born Mexican Americans and Caucasians (Hunt *et al.* 2002). They found substantially higher cardiovascular mortality rates in US-born Mexican Americans compared to Mexico-born individuals, which may be due to increased obesity and lower levels of physical exercise.

Among Native Americans, coronary disease has become the leading cause of death, and has an incidence rate almost twice that seen in Caucasian populations (Howard *et al.* 1999).

Comparisons of risk factors have shown that although Native American populations have lower total and LDL-cholesterol levels, they have higher blood pressure, higher triglyceride levels, lower HDL-cholesterol levels and a substantially greater incidence of diabetes and insulin resistance syndrome (Resnick *et al.* 2003; Zhang *et al.* 2006b). Native American groups also have higher rates of smoking compared to Caucasians, although the smoking rate has been decreasing in more recent years (Welty *et al.* 2002).

Despite these differences in risk factor clustering and rates of cardiovascular disease, a study examining whether the effects of risk factors for acute myocardial infarction vary in 52 countries comprising different ethnic groups (INTERHEART study) found that similar risk factors were seen in all populations in both men and women and in individuals of all ages (Yusuf *et al.* 2004). Nevertheless, other factors apart from orthodox risk factors such as inflammatory signalling and response may underlie differences in cardiovascular risk.

9.2.2 Markers of inflammation in groups of differing ethnicity/ancestry

Differences also exist between populations of different ethnicity/ancestry with regard to concentration of markers of inflammation, which are of current interest in relation to cardiovascular risk (see Chapter 4). Higher concentrations of fibrinogen have been found in both African American men and women compared to those of European descent (Duncan *et al.* 2000; Wong *et al.* 2006). Similarly, IL-6 concentrations were found to be higher among African American and Hispanic groups compared to Europeans in the Multi-Ethnic study of Atherosclerosis (MESA) (Wong *et al.* 2006). The same study found that levels of sICAM-1 were lower in African Americans and Asians compared to Europeans, similar to differences seen in the Wandsworth Heart and Stroke study (Miller *et al.* 2003). Differences in levels of PAI-1 also exist in different ethnic groups. The MESA study also found that levels were lower in African American individuals compared to Europeans, although higher levels were seen in Hispanic and Asian groups compared to Europeans (Wong *et al.* 2006). The IL-1Ra gene and its levels have also been studied in Caucasian and African subjects in South Africa, where differences in frequencies of IL-1Ra polymorphisms and increased levels of IL-1Ra levels were seen in African subjects compared to Caucasian subjects (Mwantembe *et al.* 2001). The factors accounting for these differences have not been explored in detail, but these may be both genetic and environmental.

9.2.3 Ethnic and ancestral differences in CRP concentration

Risk factors such as BMI and adiposity also show characteristic patterns of clustering across different ancestral backgrounds and could account in part for recent reports of differences in CRP concentration between populations. However, where adjustment has been made for some of these factors, residual population differences in CRP concentration remain that might

reflect differences in genetic regulation are observed, although residual confounding is also possible (Chambers *et al.* 2001; Forouhi *et al.* 2001; LaMonte *et al.* 2002; Chandalia *et al.* 2003).

For other factors relevant to cardiovascular risk, such as LDL- and HDL-cholesterol, it is clear that the range of regulatory and other variants in genes influencing the concentrations of these factors differ among populations. For example, the allele frequency of the –514C/T polymorphism in the hepatic lipase gene varies from 20% in Caucasians to 35% in Koreans, Chinese and Hispanics to 50% in Africans and Japanese (Jansen *et al.* 1997; Vega *et al.* 1998; Hong *et al.* 2000; Talmud *et al.* 2001; Shohet *et al.* 2002), and it has been shown that this polymorphism is consistently associated with differences in plasma HDL-cholesterol levels (Jansen *et al.* 1997; Tahvanainen *et al.* 1998; Hong *et al.* 2000). African and Japanese individuals also have higher HDL-cholesterol levels and this difference is believed to be partly due to differences in the allele frequency of the –514C/T polymorphism (Vega *et al.* 1998; Inazu *et al.* 2001; Shohet *et al.* 2002).

9.2.4 Differences in frequencies of CRP polymorphisms in different ancestral groups

Studies to investigate the potential heterogeneity of genotype-disease associations in different populations are so far limited. Those that have been undertaken are limited by the recognised inconsistency in the results of genetic association studies even within an ethnic group, due to small samples sizes. Studying the effects of genotype on an intermediate trait in populations of differing ancestry may help to examine this question with the greater power afforded by the study of an intermediate trait. Moreover, studies of variants influencing traits such as homocysteine, triglycerides, HDL- and LDL-cholesterol, and CRP indicate that genotype-intermediate phenotype associations are more tractable than gene-disease studies.

Many studies have now examined the association between genotype and CRP concentration in subjects of European descent (see Chapter 6). There is limited information on whether the same or different alleles influence CRP concentration in subjects of differing ancestry and/or whether genetic background modifies the effect size. In contrast, a number of studies have evaluated CRP concentrations in populations of differing ancestry (LaMonte *et al.* 2002; Heald *et al.* 2003; Chandalia *et al.* 2003; Ford *et al.* 2004; Lakoski *et al.* 2005; Matthews *et al.* 2005).

The CRP gene contains polymorphisms whose frequencies may differ in different populations. This might contribute in part to the observed differences in CRP concentration between groups of differing ancestry. However, there are few studies on the relationship between ancestral origin and the frequency of CRP polymorphisms. Also, no study has

addressed whether associations of CRP genotype with CRP concentration differ among subjects of differing ancestral group, i.e. whether the effects are modified by ancestry. Polymorphisms that have previously been studied are the $(GT)_n$ repeat variant in the CRP intron, the –286C/T/A and the –305G/A polymorphisms in the promoter region of the gene. Public domain resources such as the University of Washington and the Fred Hutchinson Cancer Research Center (UW-FHCRC) Variation Discovery Resource (SeattleSNPs) database (http://pga.gs.washington.edu/data/crp/) and the HapMap Consortium database (http://www.hapmap.org) have recently produced SNP maps for the CRP gene, with information on common variation in different populations that now allow assessment of these questions.

The aims of the work in this chapter were therefore to (1) quantify more precisely whether differences in CRP concentration exist between ethnic and ancestral groups; (2) to evaluate if the frequency of common genetic variants in the CRP gene differs in populations of differing ancestry; (3) to assess if differences in the frequency of CRP polymorphisms could contribute to observed differences in CRP concentration; and (4) to assess whether the CRP genotype-CRP concentration association might be modified by ethnicity/ancestry. These aims were addressed by conducting a systematic review, utilising a bioinformatic approach as well as new genotyping in a data set comprising subjects of varied ancestry from Columbia (ETNIAS study).

9.3 Methods

9.3.1. Data search from published studies

Two electronic databases (PubMed Medline and EMBASE were searched up to June 2006 for all studies examining CRP concentrations in different ancestral groups. For the search, the text words, which were also MeSH terms, "c-reactive protein" and "CRP" in combination with "ethnic", "ethnicity", "ancestry", and "ancestral" were used. The search was restricted to population based cross-sectional studies, forming a different data set to that evaluated in Chapter 6. Data from studies involving populations with the same ancestral background were pooled to obtain a summary estimate of the mean CRP concentration. Studies where an ancestral group had been compared to a Caucasian group were also pooled in a meta-analysis to obtain the mean difference in CRP concentration between ancestral groups.

9.3.2. Bioinformatics and public domain database searches

Two public domain databases were examined for information on CRP polymorphisms in different ancestral populations and are described in the Methods section. These were the University of Washington and the Fred Hutchinson Cancer Research Center (UW-FHCRC)

Variation Discovery Resource (SeattleSNPs) database (http://pga.gs.washington.edu/data/crp/) and the HapMap Consortium database (http://www.hapmap.org).

The SeattleSNPs database has re-sequenced the CRP gene using a fine-mapping technique allowing high density SNP-maps to be produced. This program has identified common variable sites in the CRP gene, and established their relative frequencies in two populations with different evolutionary histories (African and European). The African descent population consists of 24 individuals (12 male/12 female) and is composed of DNA available from the Coriell Cell Repository (http://locus.umdnj.edu/). These individuals were selected from a human variation panel of 50 African Americans. The European descent population consists of 23 individuals (12 male/11 female) and is composed of Centre d'Etude du Polymorphisme Humain (CEPH) parent DNA that are Utah residents with ancestry from northern and western Europe.

The HapMap database is a culmination of information gathered from genotyping over one million SNPs across the human genome and resulting in the production of lower density SNP maps. Two hundred and seventy individuals have been genotyped from four geographical populations. These are 30 trios (both parents and one child) of Yoruba people from Ibadan, Nigeria (YRI), 45 unrelated Japanese people from Tokyo (JPT), 45 unrelated Chinese from Beijing (CHB), and 30 CEPH trios (both parents and one child) (CEU).

9.3.3. ETNIAS study population

The ETNIAS study is a population-based cross-sectional survey, using a random sampling scheme composed of a three stage sampling process. Seven Colombian regions with different ethnic backgrounds were selected (Bucaramanga, Marinilla, Pereira, Pasto, Neiva, Quibdó and Cartagena). A total of 1028 non-related adult subjects with a White-Hispanic, Afro-Caribbean, indigenous Amerindian, or mixed ethnic background were selected. The Amerindian reservations (Tama Amerindians and Emberá Amerindians) were located in two rural areas, where inhabitants do not speak Spanish and have continued within their own cultural habits. Inclusion criteria for the study included being born and living all their life in the region they were sampled, and having previous relatives (parents and grandparents) also born in the same region. All included participants answered a survey in regards to their age, gender, socio-economic position, ethnic background, and family history of diabetes, hypertension and CVD. Blood samples were also taken at the time of the survey and were transported from each region to the central laboratory at Universidad Autonoma de Bucaramanga. DNA was extracted in Columbia from all individuals and transferred to UCL for genotyping.

9.3.4 CRP plasma measurements

Plasma CRP concentrations were measured in the ETNIAS study by immunometric assay using an Immulite Autoanalyzer (Diagnostics Products) with an analytical sensitivity lower limit of 0.1 mg/l and upper limit of 150 mg/L, and inter-assay and intra-assay coefficients of variation less than 8%.

9.3.5 Genotyping the CRP polymorphisms

Polymorphisms were chosen for genotyping in the ETNIAS study based on their ability as tagging SNPS in populations of mixed ancestry. Since African descent populations have more variation in the CRP gene compared to European descent populations, the full set of tagging SNPs used to generate common haplotypes in African descent populations (identified in Chapter 5) was used in this study, as it was expected that these SNPs would also capture most of the variation in other populations. Six polymorphisms were chosen based on data from the PHASE (v 2.0) haplotype inference program from the SeattleSNPs database (see Chapter 5).

All polymorphisms were genotyped using a standard TaqMan assay by design (Applied Biosystems), with the exception of the –286C/T/A triallelic polymorphism, which was genotyped using pyrosequencing (see Chapter 3). The -305G/A polymorphism was genotyped with the forward primer 5' GGG CTG AAG TAG GTG TTG GA 3' and the reverse primer 5' TCC TGC GAA AAT AAT GGG AAA TGG T 3'. The -717A/G polymorphism was genotyped with the forward primer 5' GCT GAG AAA ATG TGT CCA TGC AAA A 3' and the reverse primer 5' TCC TGT GTC CAA GTA TTC TCA TTG TTC 3'. The +1444C/T polymorphism was genotyped using the forward primer 5' GGT CTG GGA GCT CGT TAA CTA TG 3' and the reverse primer 5' TCC AAC TTG AAA AAC AAA ACA CCT CAA 3'. The +2302G/A polymorphism was genotyped with the forward primer 5' CAC CAG TAG CCA TCT TGT TTG C 3' and the reverse primer 5' CCA CTT CCA GTT TGG CTT CTG T 3'. The +4899T/G polymorphism was genotyped using the forward primer 5' TTA TCC TAG GAC AAC TGC CCA CTA 3' and the reverse primer 5' GGA GCT GAA GAG AAG GAA TCC A 3'. The –286C/T/A polymorphism was genotyped using a pyrosequencing method using the forward primer 5' TGA TTT GGG CTG AAG TAG GTG 3', the reverse primer 5' TGG CTA TCT ATC CTG CGA AAA T 3' and the sequencing primer 5' ACC CAG ATG GCC ACT 3'.

9.3.6 Statistical analyses

For pooled estimates of CRP concentrations among single ethnic groups from published data, meta-analyses were conducted to obtain a summary estimate for the weighted mean CRP concentration, using a random effects model to allow for any heterogeneity across studies. The CRP distribution is skewed so in some studies, CRP values were log transformed to normalise the data and studies reported either the geometric mean or the median CRP. In cases where both

the arithmetic mean and geometric mean were reported, the geometric mean was used. The majority of studies reported approximate standard deviation (SD) values, although when these values were unavailable, the SD from the largest study was applied. As the studies used differing degrees of adjustment for confounders, the unadjusted mean concentrations were used. The DerSimonian and Laird Q test, and the I^2 test (Higgins *et al.* 2003) were used to evaluate the degree of heterogeneity between studies, and funnel plots to evaluate small-study bias, of which publication bias is one potential cause. The I^2 value is the percentage of total variation across studies that is due to heterogeneity rather than chance. Data were analysed using the Comprehensive Meta-analysis (CMA) version 2.0 software (Biostat).

Statistical analyses in the ETNIAS study were conducted in Columbia under the supervision of Dr. Juan Pablo Casas. Genotype frequencies were compared to those expected under the Hardy-Weinberg equilibrium assumption by χ^2 analysis. Associations of genotype with CRP were tested using one-way analysis of variance (ANOVA) or Kruskal-Wallis tests for continuous variables and χ^2 or Fisher's exact test for categorical variables.

9.4 Results

9.4.1 Published studies on CRP concentration in groups of differing ancestry

Eighteen publications that fulfilled the inclusion criteria were identified with data on CRP concentrations (see Table 9.1). All studies had information on absolute values of CRP in a single ancestral group (Forouhi *et al.* 2001; Yamada *et al.* 2001; Chatha *et al.* 2002a; LaMonte *et al.* 2002; Chandalia *et al.* 2003; Lear *et al.* 2003; Ford *et al.* 2003b; Ford *et al.* 2003a; Heald *et al.* 2003; Ford *et al.* 2004; Albert *et al.* 2004; Anand *et al.* 2004; Sinha *et al.* 2005; Khera *et al.* 2005; Matthews *et al.* 2005; Lakoski *et al.* 2005; Bravata *et al.* 2005; Diez Roux *et al.* 2006). Sixteen studies had information on both absolute values of CRP and differences in concentration between Caucasian and non-Caucasian populations (Forouhi *et al.* 2001; Chatha *et al.* 2002b; LaMonte *et al.* 2002; Chandalia *et al.* 2003; Lear *et al.* 2003; Ford *et al.* 2003b; Ford *et al.* 2003a; Heald *et al.* 2003; Ford *et al.* 2004; Albert *et al.* 2004; Anand *et al.* 2004; Khera *et al.* 2005; Matthews *et al.* 2005; Lakoski *et al.* 2005; Bravata *et al.* 2005; Diez Roux *et al.* 2006). Two additional publications were also identified but were subsequently excluded for reasons detailed below.

Categorisation of ancestral background utilised study-specific definitions and identified five major ancestral groupings. These were African American, Caucasian, South Asian (Indian or Pakistani origin), East Asian (Chinese or Japanese origin) and Hispanic (Hispanic and Mexican American). Within studies, ancestral background was usually self-reported though in many cases, information on the parents' background was also obtained. The study by Heald et

al. was the only study to report data on Africans of Caribbean origin and was therefore excluded from the comparison between Caucasians and African Americans in order to keep this group as homogeneous as possible (Heald *et al.* 2003). The study by Bravata et al. was also excluded from the analyses as the CRP measures were taken after an incident of stroke and were therefore not comparable to the other studies that measured baseline plasma CRP in healthy subjects.

Of the studies where absolute values of CRP in a single population were reported, eleven fell into the African American group (11362 individuals), sixteen in the Caucasian group (32751 individuals), six in the South Asian group (784 individuals), seven in the East Asian group (8965 individuals) and seven in the Hispanic group (6033 individuals). Of studies where the CRP concentration had been compared between Caucasian and non-Caucasian groups, eleven of the studies included data on African American subjects, five on South Asian subjects, six on East Asian subjects, and seven on Hispanic subjects. Studies that had reported the CRP concentration in both men and women separately were entered as two data sets.

Figure 9.2. Flowchart to show publications included on CRP concentration in different ancestral groups.

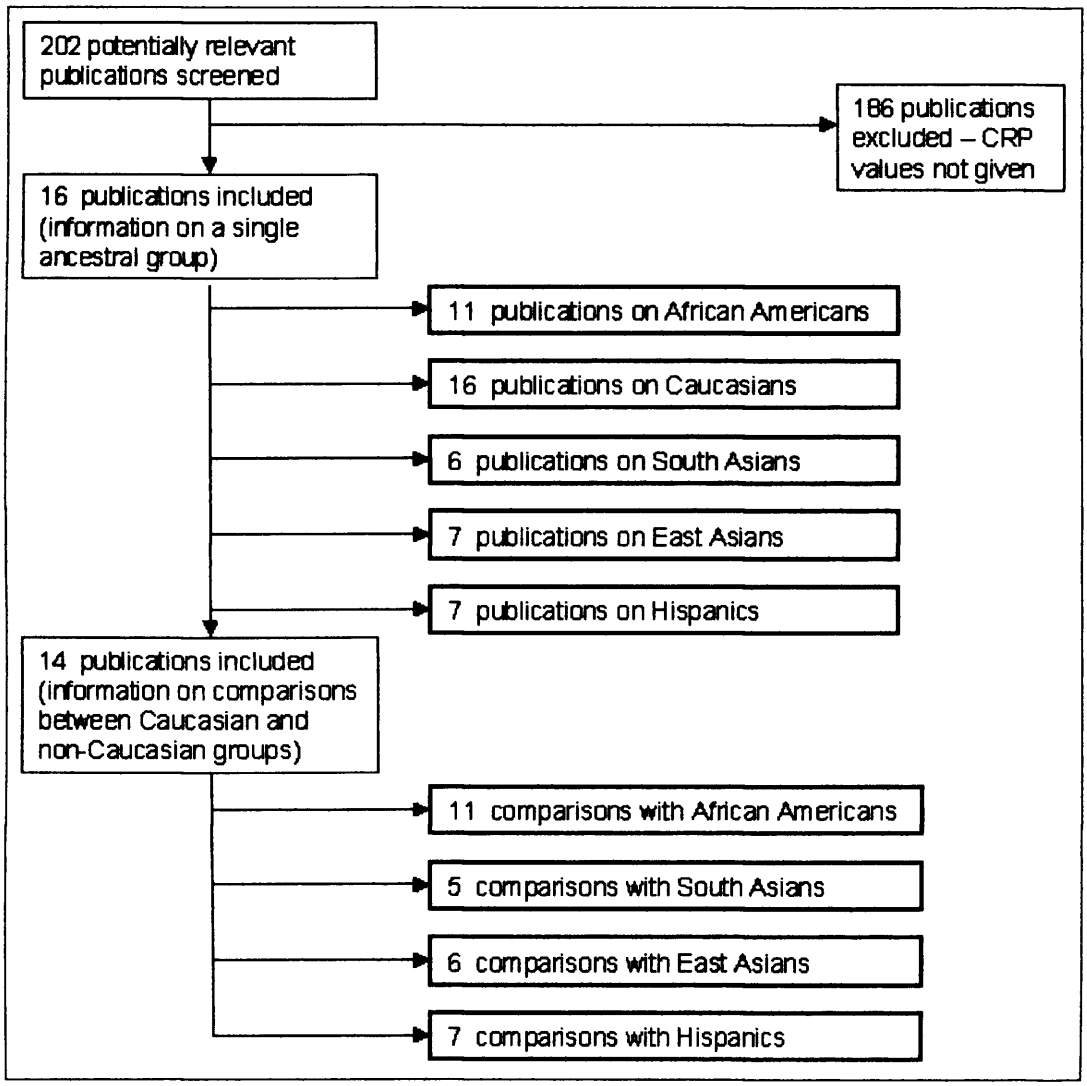


Table 9.1. Characteristics of published studies on CRP concentration.

Study	Year published	CRP in single population	CRP differences between populations	Population(s) studied	CRP m
Forouhi	2001	Yes	Yes	Cauc; SAs	E
Yamada	2001	Yes	No	As	Immunor
Chatha	2002	Yes	Yes	Cauc; SAs	Imm
LaMonte (CAPS)	2002	Yes	Yes	AfAm; Cauc	Immunor
Chandalia	2003	Yes	Yes	Cauc; SAs	Immunor
Lear	2003	Yes	Yes	Cauc; As	Chemi-ill
Ford (NHANES <20 years)	2003	Yes	Yes	AfAm; Cauc	Immunor
Ford (NHANES men)	2003	Yes	Yes	AfAm; Cauc	Immunor
Heald	2003	Yes	Yes	AfAm; Cauc; SAs	E
Ford (NHANES women)	2004	Yes	Yes	AfAm; Cauc	Immunor
Albert (WHS)	2004	Yes	Yes	AfAm; Cauc; As; His	Immunor
Anand	2004	Yes	Yes	Cauc; SAs; As	Immunor
Sinha	2005	Yes	No	SAs	E
Khera (Dallas)	2005	Yes	Yes	AfAm; Cauc	Immunot
Matthews (SWAN)	2005	Yes	Yes	AfAm; Cauc; As; His	Immunor
Lakoski (MESA)	2005	Yes	Yes	AfAm; Cauc; As; His	Immunor
Bravata (NHANES)	2005	Yes	Yes	AfAm; Cauc	Immunor
Diez Roux (MESA)	2006	Yes	Yes	AfAm; Cauc; As; His	Immunor

AfAm = African American, Cauc = Caucasian, SAs = South Asian, As = Asian and His = Hispanic.

ELISA = Enzyme Linked Immunosorbent Assay.

9.4.2 CRP concentrations according to ancestral background

The mean CRP concentrations were pooled from each study by ancestral group to obtain a summary estimate of CRP concentration within ethnic groups, although mean CRP concentrations within a group varied substantially (see Tables 9.9-9.6). Estimates of mean CRP concentrations suggested that African American populations have the highest CRP concentrations (2.83mg/L, 95%CI: 2.06-3.61) and Asian populations including Chinese and Japanese groups have the lowest CRP concentrations (0.84mg/L, 95%CI: 0.59-1.08) (see Figure 9.3).

Table 9.2. Measures in the studies with data on CRP concentration in African Americans.

Study	N	Geometric mean CRP (mg/L) (SD)	Geometric mean age (years)	Geometric mean BMI (kg/m^2)	Geometric mean sBP (mmHg)	Current smokers (%)
La Monte (CAPS)	44	4.3 (1.99)	56.6	30.9	129	8.7
Ford (NHANES <20 years)	963	1.7 (3.1)	-	-	-	-
Ford (NHANES men)	352	1.7 (3.0)	-	-	-	-
Heald (men)*	82	1.0 (1.39)†	52.4	27.7	-	-
Heald (women)*	111	1.3 (1.61)†	47.5	28.8	-	-
Ford (NHANES women)	419	3.1 (2.05)	-	-	-	-
Albert (WHS)	475	2.96 (3.46)†	54.0	28.6	-	16.2
Khera (Dallas men)	740	2.0 (4.0)	43.7	29.6	-	32
Khera (Dallas women)	1018	3.2 (4.2)	43.5	33.3	-	26.3
Matthews (SWAN)	729	3.0 (4.6)†	46.3	31.2	119.0	24.3
Lakoski (MESA)	1871	2.07 (7.3)	-	-	-	-
Bravata (NHANES)	2752	7.0 (10.0)	58.2	28.4	135.3	32.0
Diex Roux (MESA)	1806	1.91 (4.97)	-	-	-	-

*African Caribbean.

†Median CRP values reported and approximate SD calculated from the interquartile range.

Table 9.3. Measures in the studies with data on CRP concentration in Caucasians.

Study	N	Geometric mean CRP (mg/L) (SD)	Geometric mean age (years)	Geometric mean BMI (kg/m^2)	Geometric mean sBP (mmHg)	Current smokers (%)
Forouhi (men)	28	0.92 (0.94)†	-	26.1	-	64
Forouhi (women)	29	0.70 (0.96)†	-	24.9	-	59
Chatha (men)	63	1.77 (1.45)	55.2	-	139.8	-
Chatha (women)	58	2.23 (1.54)	56.1	-	136.2	-
La Monte (CAPS)	46	2.3 (8.8)	54.3	25.2	116.3	6.5
Chandalia	55	0.63 (4.2)	29	26	120	-
Lear (men)	39	0.63 (3.7)†	38.9	25.6	-	-
Lear (women)	52	0.72 (3.7)†	41	27.1	-	-
Ford (NHANES <20 years)	713	1.4 (2.64)	-	-	-	-
Ford (NHANES men)	911	1.6 (2.8)	-	-	-	-
Heald (men)	72	2.2 (1.51)†	51.4	27.2	-	-
Heald (women)	70	2.1 (1.79)†	52.4	27.0	-	-
Ford (NHANES women)	963	2.3 (1.84)	-	-	-	-
Albert (WHS)	24455	2.02 (2.64)†	54.7	25.9	-	11.5
Anand	322	2.49 (3.7)	51.3	27.5	118.9	-
Khera (Dallas men)	475	1.8 (3.1)	45.1	28.9	-	24.7
Khera (Dallas women)	516	2.6 (3.7)	46.8	29.0	-	26.2
Matthews (SWAN)	1400	1.4 (2.45)†	46.4	27.4	111.0	15.4
Lakoski (MESA)	2602	1.55 (4.59)	-	-	-	-
Bravata (NHANES)	8411	5.0 (8.9)	64.1	27.2	134.0	18.0
Diex Roux (MESA)	2746	1.84 (2.43)	-	-	-	-

†Median CRP values reported and approximate SD calculated from the interquartile range.

Table 9.4. Measures in the studies with data on CRP concentration in South Asians.

Study	N	Geometric mean CRP (mg/L) (SD)	Geometric mean age (years)	Geometric mean BMI (kg/m^2)	Geometric mean sBP (mmHg)	Current smokers (%)
Forouhi (men)	28	1.07 (0.55)†	-	25.7	-	21
Forouhi (women)	31	1.35 (1.72)†	-	25.5	-	0
Chatha (men)	39	1.94 (1.46)	52.1	-	136.2	-
Chatha (women)	31	2.29 (1.52)	53.3	-	136.5	-
Chandalia	82	0.94 (4.2)	31	23.9	112	-
Heald (men)	68	1.7 (2.1)†	50.4	27.2	-	-
Heald (women)	62	2.8 (3.2)†	49.6	29.2	-	-
Anand	323	3.22 (4.2)	49.4	26.3	119.7	-
Sinha	30	1.3 (1.85)	40.6	23.2	-	-

†Median CRP values reported and approximate SD calculated from the interquartile range.

Table 9.5. Measures in the studies with data on CRP concentration in East Asians.

Study	N	Geometric mean CRP (mg/L) (SD)	Geometric mean age (years)	Geometric mean BMI (kg/m^2)	Geometric mean sBP (mmHg)	Current smokers (%)
Yamada (men)	2275	0.83 (3.6)	55.5	22.7	129.4	50.1
Yamada (women)	3832	0.59 (2.7)	55.9	23.0	127.1	6.5
Lear (men)	44	0.36 (1.31)†	34.8	23.4	-	-
Lear (women)	47	0.33 (1.30)†	41.6	23.1	-	-
Albert (WHS)	357	1.12 (1.31)†	53.3	23.4	-	3.3
Anand	306	1.50 (3.3)	47.7	23.9	119.2	-
Matthews (SWAN Chinese)	231	0.7 (0.82)†	46.6	23.2	108.5	1.7
Matthews (SWAN Japanese)	248	0.5 (0.74)	46.7	23.0	109	12.1
Lakoski (MESA)	801	0.82 (2.26)	-	-	-	-
Diez Roux (MESA)	824	1.32 (2.52)	-	-	-	-

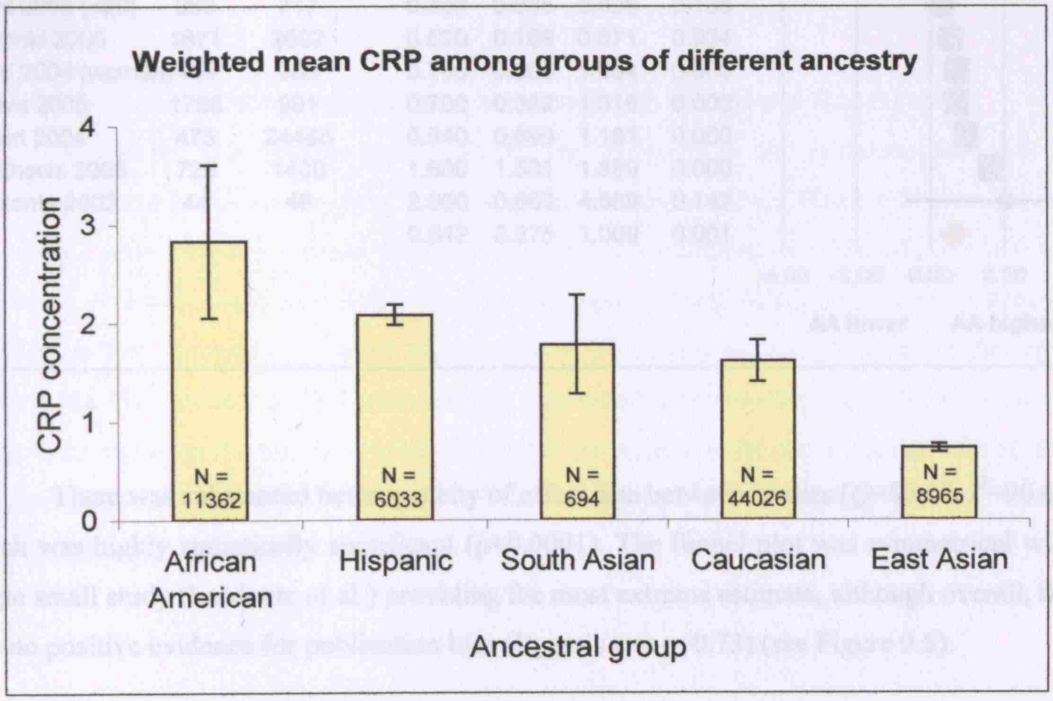
†Median CRP values reported and approximate SD calculated from the interquartile range.

Table 9.6. Measures in the studies with data on CRP concentration in Hispanics.

Study	N	Geometric mean CRP (mg/L) (SD)	Geometric mean age (years)	Geometric mean BMI (kg/m^2)	Geometric mean sBP (mmHg)	Current smokers (%)
Ford (NHANES <20 years old)	1433	2.0 (2.96)	-	-	-	-
Ford (NHANES men)	516	1.7 (2.8)	-	-	-	-
Ford (NHANES women)	618	3.2 (2.4)	-	-	-	-
Albert (WHS)	254	2.06 (2.95)†	54.0	26.1	-	10.2
Matthews (SWAN)	226	2.3 (3.04)†	46.2	28.8	120.0	17.7
Lakoski (MESA)	1488	2.24 (5.79)	-	-	-	-
Diez Roux (MESA)	438	1.58 (3.51)	-	-	-	-

†Median CRP values reported and approximate SD calculated from the interquartile range.

Figure 9.3. Mean CRP concentration and 95% confidence interval among different ancestral groups.

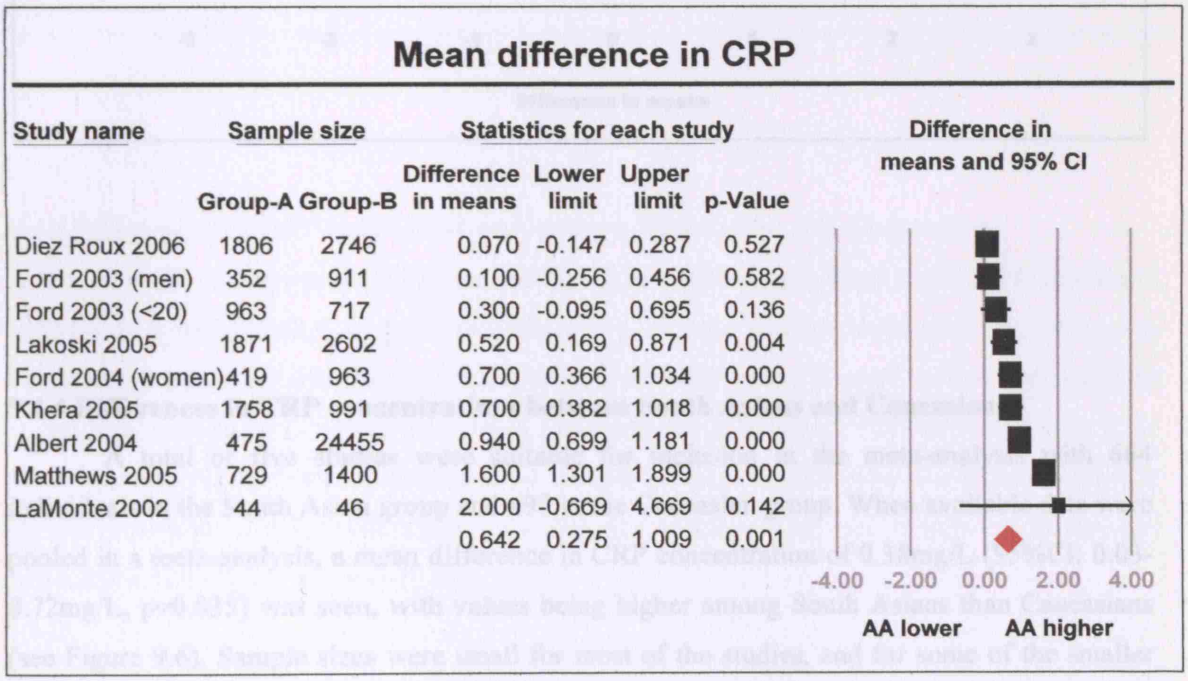


In studies where a group of different ancestral background was compared with a Caucasian group, the weighted mean difference (WMD) in CRP concentration was calculated and the data pooled in a meta-analysis. A random effects model was used to allow for heterogeneity between the studies. For all studies, the reported unadjusted mean CRP concentration was used to determine the difference between ancestral groups.

9.4.3 Differences in CRP concentrations between African Americans and Caucasians

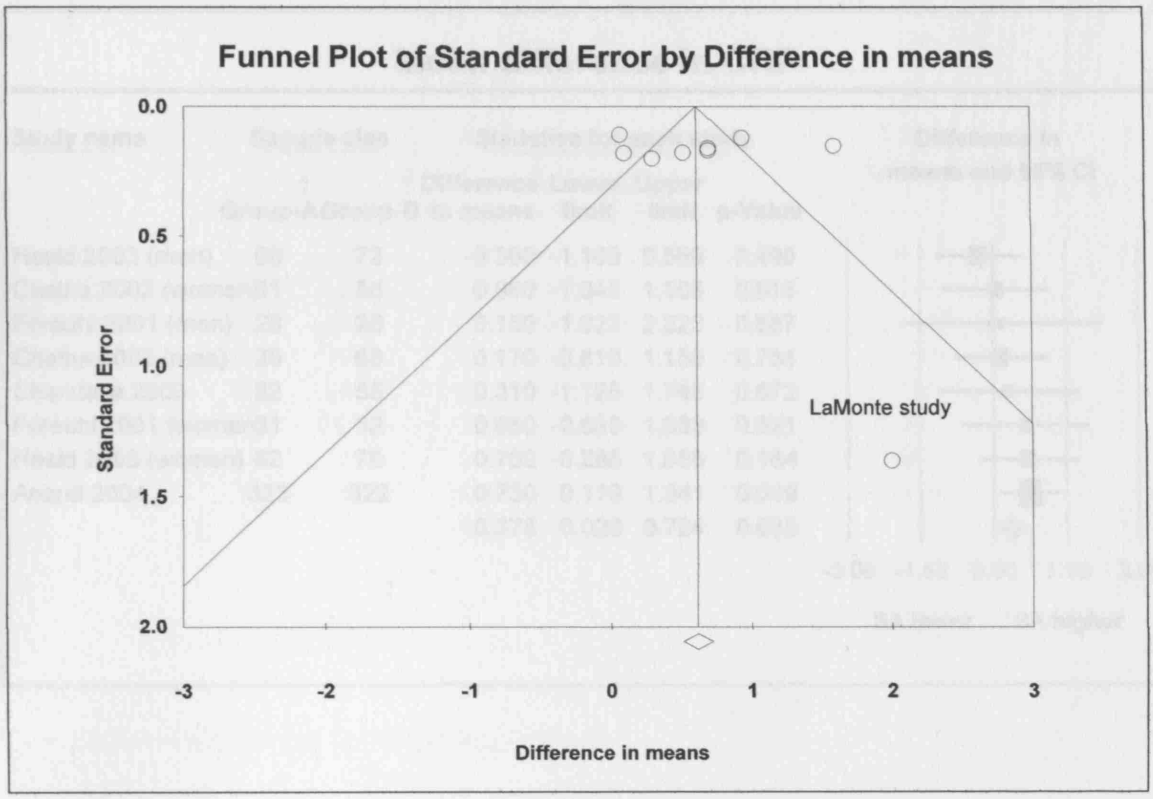
Data on the mean difference in CRP concentration between African Americans and Caucasians was available from nine studies comprising 8417 African American individuals and 34831 Caucasian individuals (see Figure 9.4). The CRP concentration was 0.64mg/L (95%CI: 0.28-1.01mg/L) higher among African American subjects than Caucasian subjects, which was highly significant (p=0.001).

Figure 9.4. Mean difference in CRP between African Americans and Caucasians.



There was substantial heterogeneity of effect size between studies (Q=85.09, I^2=90.6%), which was highly statistically significant (p<0.0001). The funnel plot was symmetrical with a single small study (LaMonte et al.) providing the most extreme estimate, although overall, there was no positive evidence for publication bias (Egger's test p=0.73) (see Figure 9.5).

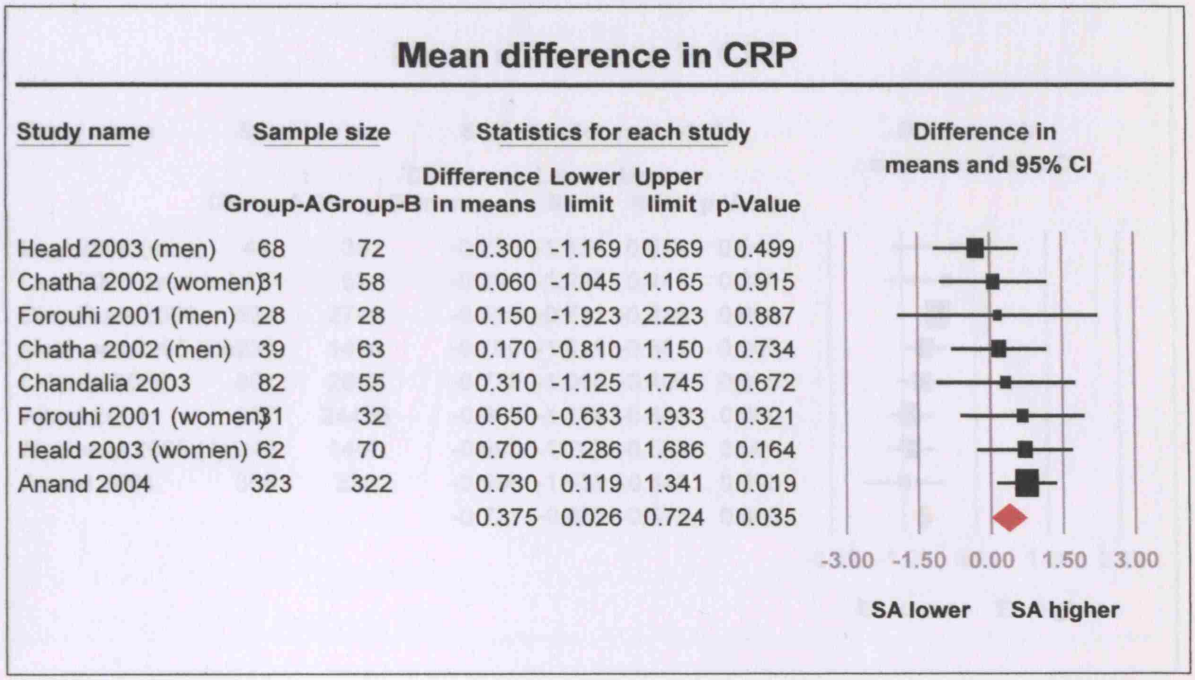
Figure 9.5. Funnel plot to show presence of bias in the studies used in the meta-analysis.



9.4.4 Differences in CRP concentrations between South Asians and Caucasians

A total of five studies were suitable for inclusion in the meta-analysis with 664 individuals in the South Asian group and 697 in the Caucasian group. When available data were pooled in a meta-analysis, a mean difference in CRP concentration of 0.38mg/L (95%CI: 0.03-0.72mg/L, p=0.035) was seen, with values being higher among South Asians than Caucasians (see Figure 9.6). Sample sizes were small for most of the studies, and for some of the smaller studies, the SD values reported were implausibly small and therefore the SD value for each group from the largest study was used. There did not appear to be any heterogeneity of effect size between studies (Q=4.74, p=0.69, I^2=0.0%). The Egger's test provided no positive evidence for publication bias (p=0.47).

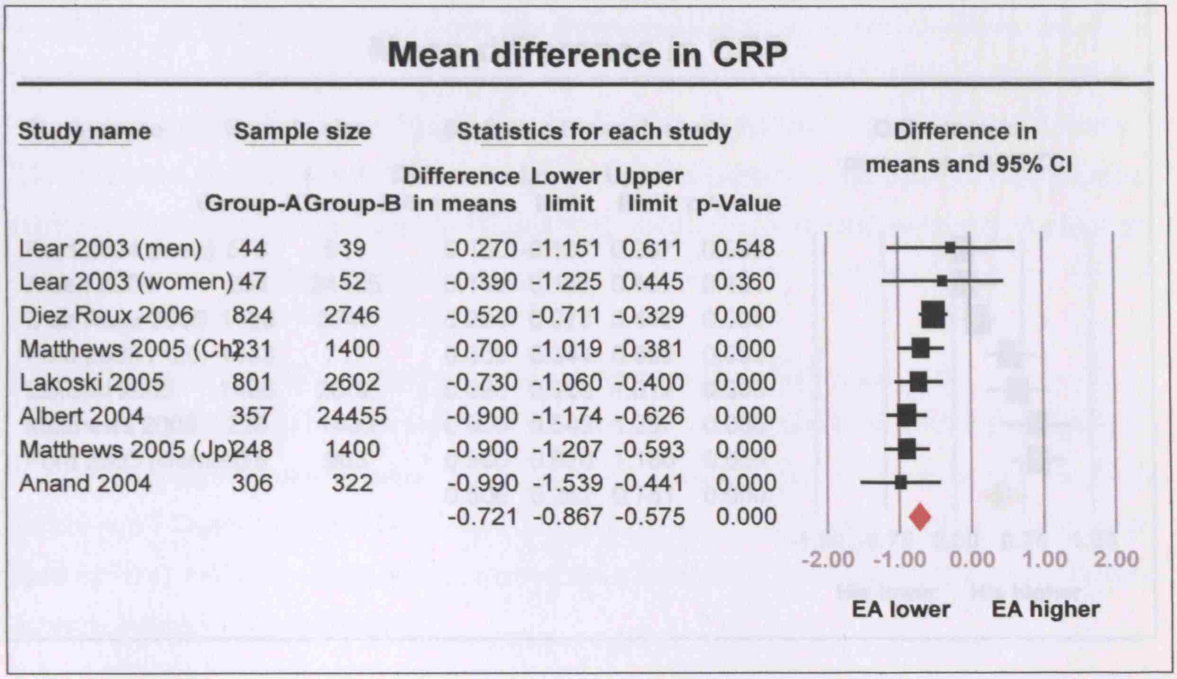
Figure 9.6. Mean difference in CRP between South Asians and Caucasians



9.4.5 Differences in CRP concentrations between East Asians and Caucasians

Data from six studies with a total of 2858 East Asian individuals and 33016 Caucasian individuals were pooled. The mean difference in CRP concentration between East Asian and Caucasian populations was found to be 0.72mg/L (95%CI: 0.58-0.87mg/L) (lower among East Asians than Caucasians), which was highly statistically significant ($p<0.0001$) (see Figure 9.7). When the differences between studies were examined, there was no evidence of heterogeneity (Q=9.63, p=0.21; I^2=27.33%). There also did not appear to be any positive evidence for publication bias (Egger's test, p=0.83).

Figure 9.7. Mean difference in CRP between East Asians and Caucasians



Ch = Chinese subjects and Jp = Japanese subjects.

9.4.6 Differences in CRP concentrations between Hispanics and Caucasians

CRP concentrations have also been compared between Hispanic and Caucasian populations in seven large studies, allowing the data to be pooled in a meta-analysis (see Figure 9.8). Individual studies were generally large in size, comprising a total of 6033 Hispanic individuals and 33794 Caucasian individuals. The mean difference in CRP concentration was 0.51mg/L (95%CI: 0.16-0.81mg/L), with values being higher among Hispanics than Caucasians, which was highly significant (p=0.004). There was substantial heterogeneity of effect size between studies (Q=32.43, I^2=81.5%), which was highly statistically significant (p<0.001), although overall, there was no positive evidence for publication bias (Egger's test p=0.37) (see Figure 9.9).

Figure 9.8. Mean difference in CRP between Hispanics and Caucasians

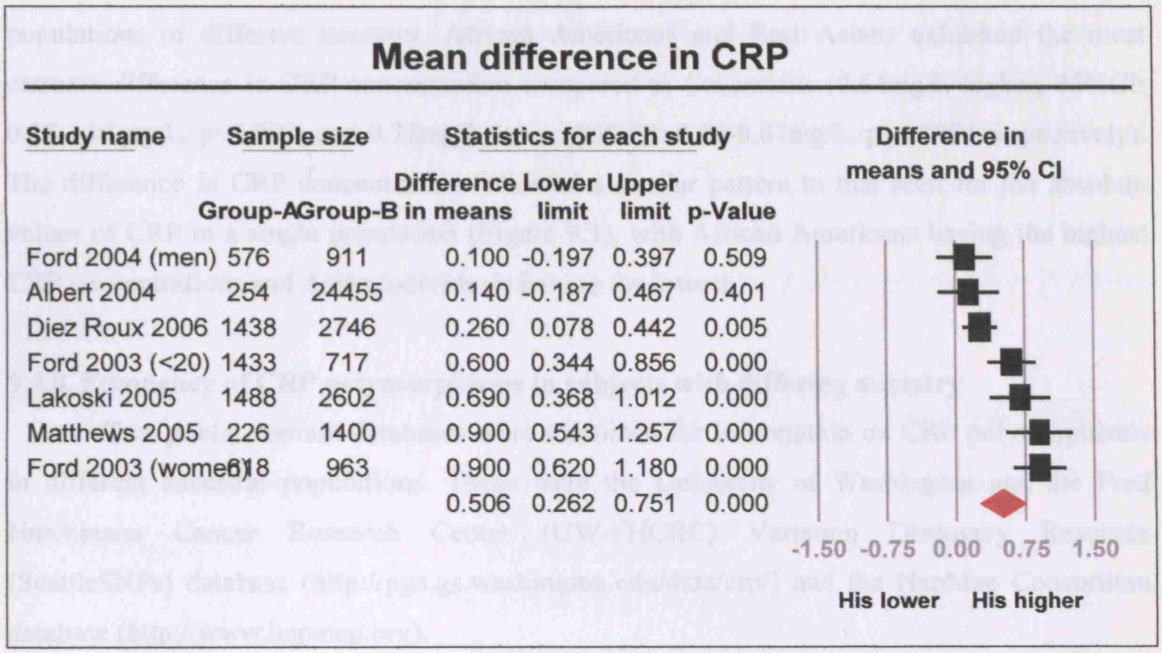
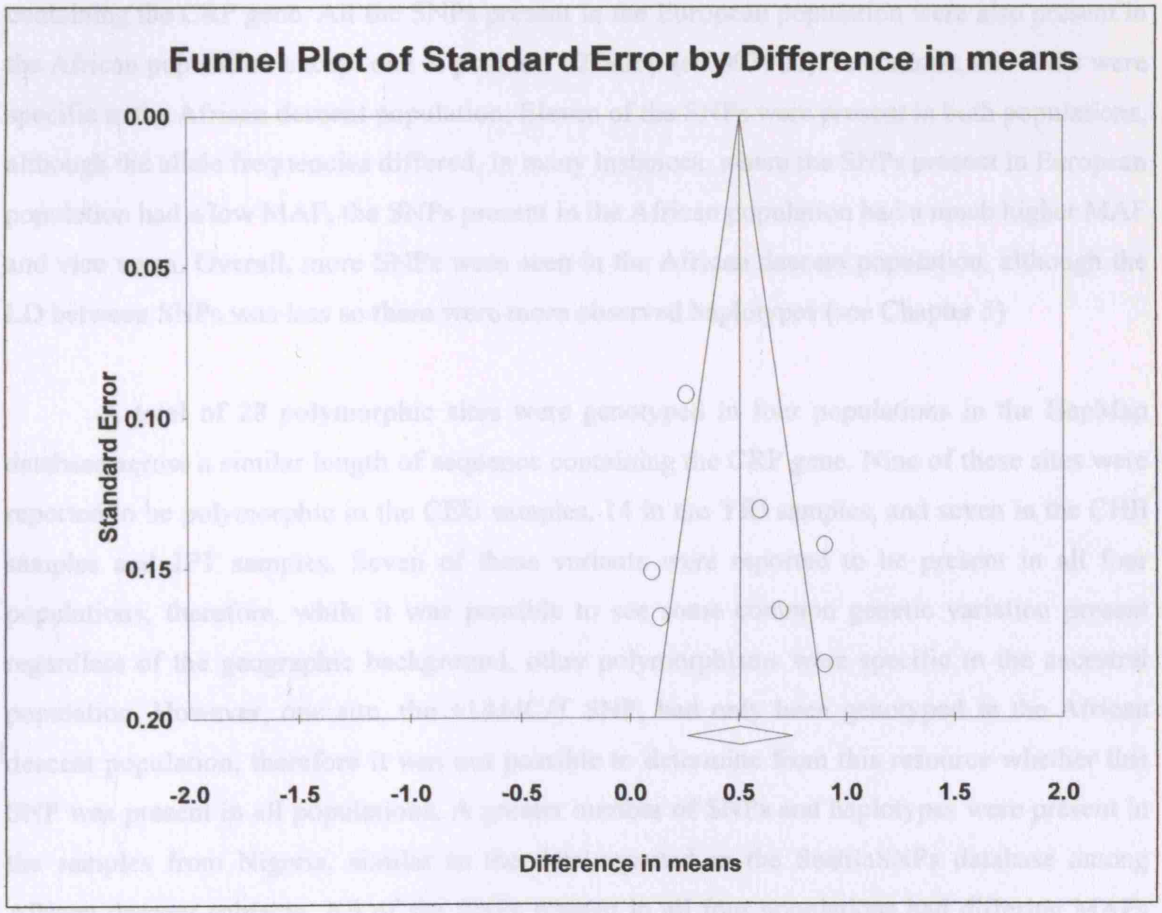


Figure 9.9. Funnel plot to show presence of bias in the studies used in the meta-analysis.



9.4.7. Overview of results from the systematic review

Overall, substantial differences were found to exist in CRP concentration between populations of different ancestry. African Americans and East Asians exhibited the most extreme difference in CRP concentration compared to Caucasians (0.64mg/L higher, 95%CI: 0.28-1.01mg/L, p=0.001; and 0.72mg/L lower, 95%CI: 0.58-0.87mg/L, p<0.0001 respectively). The difference in CRP concentration followed a similar pattern to that seen for the absolute values of CRP in a single population (Figure 9.3), with African Americans having the highest CRP concentrations and Asian individuals having the lowest.

9.4.8. Frequency of CRP polymorphisms in subjects with differing ancestry

Two public domain databases were examined for information on CRP polymorphisms in different ancestral populations. These were the University of Washington and the Fred Hutchinson Cancer Research Center (UW-FHCRC) Variation Discovery Resource (SeattleSNPs) database (http://pga.gs.washington.edu/data/crp/) and the HapMap Consortium database (http://www.hapmap.org).

The SeattleSNPs database identified 12 common polymorphisms with a minor allele frequency of 0.05 or more in the CRP gene in the population of European descent and 29 common polymorphisms in the population of African descent across a region of around 6050bp containing the CRP gene. All the SNPs present in the European population were also present in the African population except one at position +2792bp (rs3093080). In contrast, ten SNPs were specific to the African descent population. Eleven of the SNPs were present in both populations, although the allele frequencies differed. In many instances, where the SNPs present in European population had a low MAF, the SNPs present in the African population had a much higher MAF and vice versa. Overall, more SNPs were seen in the African descent population, although the LD between SNPs was less so there were more observed haplotypes (see Chapter 5)

A total of 28 polymorphic sites were genotyped in four populations in the HapMap database across a similar length of sequence containing the CRP gene. Nine of these sites were reported to be polymorphic in the CEU samples, 14 in the YRI samples, and seven in the CHB samples and JPT samples. Seven of these variants were reported to be present in all four populations; therefore, while it was possible to see some common genetic variation present regardless of the geographic background, other polymorphisms were specific to the ancestral population. However, one site, the +1444C/T SNP, had only been genotyped in the African descent population, therefore it was not possible to determine from this resource whether this SNP was present in all populations. A greater number of SNPs and haplotypes were present in the samples from Nigeria, similar to the data reported in the SeattleSNPs database among African descent subjects. All of the SNPs present in all four populations had differing MAFs

according to the sample population and some SNPs had a reversal of the common and rare alleles, such as the +2302G/A variant (rs1205), where the A-allele was the common allele in the Chinese and Japanese samples and the rare allele in the European and African samples (see Table 9.8). In addition, the European samples exhibited greater LD between SNPs and therefore fewer haplotypes, and African samples had less LD between SNPs and the largest number of haplotypes.

Table 9.7. CRP polymorphisms specific and common to subjects of European descent and African descent from the SeattleSNPs and HapMap databases.

Specific to European descent subjects	Common to European and African descent subjects	Specific to African descent subjects
rs3093080 (+2792T/A)	rs3093059 (-757T/C)	rs3093058 (-936A/T)
	rs2794521 (-717A/G)	rs3093061 (-603A/G)
	rs3091244 (-286C/T/A)	rs3093062 (-305G/A)
	rs1417938 (+194A/T)	rs3093064 (-2C/T)
	rs1800947 (+1059G/C)	rs3093066 (+1436C/A)
	rs1130864 (+1444C/T)	rs3093067 (+2007T/C)
	rs1205 (+2302G/A)	rs2808631 (+2489A/G)
	rs3093068 (+3171C/G)	rs3093069 (+3404T/G)
	rs2808630 (+3667A/G)	rs3093074 (+/-T)
	rs3093075 (+4622C/A)	rs3093076 (+/-GAT)
	rs3093077 (+4899T/G)	

No ethnic specific SNPs were found in the Chinese and Japanese populations.
Minor allele frequencies of polymorphisms are greater than 2%.

Table 9.8. Minor allele frequencies of CRP SNPs common to all populations from the SeattleSNPs (http://pga.gs.washington.edu/data/crp) and HapMap (http;//www.hapmap.org) databases.

SNP id	Location	Alleles	AD freq	ED freq	Ch freq	Jp freq
rs3093059	-757b	T/C	0.35	0.07	0.12	0.11
rs1417938	+194b	A/T	0.10	0.26	0.03	0.05
rs1800947	+1059b	G/C	0.02	0.07	0.07	0.04
rs1205	+2302b	G/A	0.14	0.30	0.56	0.73
rs2808630	+3667b	A/G	0.06	0.36	0.24	0.11
rs3093075	+4622b	C/A	0.34	0.07	0.13	0.10
rs3093077	+4899b	T/G	0.39	0.07	0.14	0.10

AD: African Descent; ED: European Descent; Ch: Chinese Descent; Jp: Japanese Descent; freq: allele frequency.

9.4.9. Identification of rare polymorphisms in the CRP gene

Although public domain resources provide comprehensive data about common genetic variants, they are less informative for rare variants. The CRP gene has recently been resequenced in a greater number of subjects in order to identify polymorphisms with a low frequency (MAF<0.02). Approximately 1000 chromosomes were sequenced using subjects from the SeattleSNPs database (African American and European samples), along with Chinese, Japanese, Southeast Asian, Mexican and Indo-Pakistani subjects from an extended DNA panel, and the polymorphism discovery resource panel (PDR) (Crawford *et al.* 2006). Twenty-nine SNPs were identified in African American samples, seventeen in the Asian samples, fourteen in the Mexican samples, twelve in the European samples, and eleven in the Indo-Pakistani samples.

Sequencing of the extended panel and PDR panel identified nine new polymorphisms that were not previously identified in the SeattleSNPs database. Of these SNPs, three were specific to Chinese and Japanese subjects, one of which occurred in exon 2 (+867C/T, MAF=0.14) but did not result in an amino acid change. Two of these SNPs were only present in the Indo-Pakistani samples, one of which was a SNP in the promoter region (-752G/A, MAF=0.06), and the other was a nonsynonymous SNP in exon 2 (+905C/T, MAF=0.06) that resulted in an amino acid change from proline to lysine but was only seen in one subject. In addition, two SNPs were also specific to the Mexican subjects, one in the upstream promoter region (-948A/T, MAF=0.05), and one in the 3'UTR (+1865T/C, MAF=0.05). Allele frequencies of SNPS that were common to more than two populations are shown in Table 9.9.

Table 9.9. Minor allele frequencies of CRP SNPs common to more than two populations examined in the deep resequencing work (Crawford *et al.* 2006).

SNP id	Location	Alleles	AD freq	ED freq	Ch freq	Jp freq	SAs freq	Mx freq	IP freq
rs3093059	-757b	T/C	0.35	0.07	0.14	0.14	0.25	0.00	0.13
rs2794521	-717b	A/G	0.07	0.35	0.14	0.00	0.15	0.40	0.13
rs3093062	-305b	G/A	0.24	0.00	0.00	0.00	0.00	0.10	0.00
rs3091244	-286b	C/T	0.26	0.62	NG	NG	0.06	0.25	0.00
		C/A	0.33	0.05	NG	NG	0.19	0.00	0.17
rs1417938	+194b	A/T	0.14	0.27	0.00	0.07	0.10	0.15	0.00
rs1800947	+1059b	G/C	0.02	0.05	0.14	0.00	0.05	0.00	0.13
rs1130864	+1444b	C/T	0.14	0.32	0.00	0.00	0.10	0.15	0.00
rs1205	+2302b	G/A	0.14	0.25	0.71	0.79	0.50	0.30	0.44
rs3093068	+3171b	C/G	0.28	0.07	0.10	0.14	0.25	0.00	0.13
rs2828630	+3667b	A/G	0.06	0.36	0.14	0.00	0.15	0.40	0.13
rs3093075	+4622b	C/A	0.33	0.07	0.14	0.14	0.22	0.00	0.19
rs3093077	+4899b	T/G	0.37	0.07	0.14	0.14	0.10	0.00	0.19

AD: African Descent; ED: European Descent; Ch: Chinese Descent; Jp: Japanese Descent; SAs: Southeast Asian Descent; Mx: Mexican Descent; IP: Indo-Pakistani Descent; freq: minor allele frequency; NG: not given.

Based on published work on individual CRP variants, new resequencing work and data available from the SeattleSNPs and HapMap databases, it was possible to see that substantial differences exist at polymorphic sites along the CRP gene in different populations. Many of the genetic variants were restricted to a particular population, and while others occurred in several populations, they had different minor allele frequencies. In some cases, there was also a reversal of the minor allele, which became the common allele in some populations.

9.4.10 CRP concentrations in the ETNIAS study

In order to assess if differences in the frequency of CRP polymorphisms could contribute to differences in CRP concentration in different ancestral groups, new genotyping was conducted in the ETNIAS study. A total of 1033 individuals from four different ancestral backgrounds were recruited from nine centres in Columbia (see Table 9.10). These were White-Hispanics (n=401), Amerindians (n=63), African descent (n=145), and those with a mixed ancestral origin (n=424).

Table 9.10. Ancestral background by site of recruitment.

Site of recruitment	Ancestral background				Total number
	White-Hispanic	Amerindian	African Descent	Mixed	
Bucaramanga	135	-	-	3	138
Cartagena	8	-	17	112	137
Marinilla	76	-	-	63	139
Neiva	42	-	1	97	140
Pasto	54	4	1	82	141
Pereira	86	-	4	51	141
Quibdó	-	1	122	16	139
Emberá	-	28	-	-	28
Tamá	-	30	-	-	30
Total number	401	63	145	424	1033

Differences in socio-economic position and paternal/maternal history of clinical measures such as diabetes, hypertension and angina were seen between the ancestral groups (see Table 9.11). CRP concentrations according to demographic and clinical measures were then examined. Overall, CRP concentrations ranged between 0.05mg/L and 65mg/L, and varied significantly by gender, age group, socio-economic position and paternal/maternal history of hypertension (see Table 9.12).

Table 9.11. Demographics and other characteristics of participants in the ETNIAS study according to ancestral group.

Variable	Ethnic background				p value*
	White-Hispanic	Amerindian	African Descent	Mixed	
Age group (years)					
18-30	203 (50.6%)	37 (58.7%)	69 (47.6%)	221 (52.7%)	0.351
31-40	129 (32.2%)	12 (19.1%)	44 (30.4%)	127 (30.3%)	
41-50	69 (17.2%)	14 (22.2%)	32 (22.1%)	71 (17.0%)	
Gender					
Male	172 (42.9%)	34 (54.0%)	71 (49.0%)	195 (46.5%)	0.292
Female	229 (57.1%)	29 (46.0%)	74 (51.0%)	224 (53.5%)	
Socio-economic position					
Low	176 (44.0%)	33 (100%)	136 (94.4%)	266 (63.6%)	<0.001
Middle	209 (52.3%)	0	8 (5.6%)	147 (35.2%)	
High	15 (3.8%)	0	0	5 (1.2%)	
Paternal/maternal history of diabetes					
None	229 (86.9%)	62 (98.4%)	129 (89.0%)	64 (15.3%)	0.021
Father, mother or both	37 (13.1%)	1 (1.6%)	16 (11.0%)	354 (84.7%)	
Paternal/maternal history of hypertension					
None	149 (56.0%)	57 (90.5%)	77 (53.1%)	228 (54.6%)	<0.001
Father or mother	104 (39.1%)	6 (9.5%)	54 (37.2%)	160 (38.3%)	
Father and mother	13 (4.9%)	0	14 (9.7%)	30 (7.2%)	
Paternal/maternal history of angina					
None	237(89.1%)	63 (100%)	138 (95.2%)	369 (88.3%)	0.004
Father, mother or both	29 (10.9%)	0	7 (4.8%)	49 (11.7%)	
Paternal/maternal history of MI					
None	250 (94.0%)	63 (100%)	137 (94.5%)	393 (94.0%)	0.262
Father, mother or both	16 (6.0%)	0	8 (5.5%)	25 (6.0%)	
Paternal/maternal history of stroke					
None	255 (95.9%)	62 (98.4%)	137 (94.5%)	391 (93.5%)	0.308
Father, mother or both	11 (4.1%)	1 (1.6%)	8 (5.5%)	27 (6.5%)	

*χ^2 analysis.

Table 9.12. CRP concentration by city of recruitment, ancestral background and other demographic characteristics.

Variable	Number of subjects	Geometric mean CRP (mg/L) (95% confidence interval)	Kruskal-Wallis test*
City of recruitment			
Bucaramanga	58	0.93 (0.62-1.39)	$\chi^2 = 16.926$
Cartagena	134	1.55 (1.25-1.92)	$p = 0.032$
Marinilla	138	1.24 (1.00-1.54)	
Neiva	137	1.47 (1.16-1.86)	
Pasto	133	1.34 (1.09-1.66)	
Pereira	136	1.22 (0.99-1.50)	
Quibdo	139	0.87 (0.69-1.11)	
Emberá	28	0.80 (0.43-0.46)	
Tamá	28	0.91 (0.17-0.80)	
Ancestral background			
White-Hispanic	318	1.21 (1.05-1.39)	$\chi^2 = 7.578$
Amerindian	61	0.92 (0.61-1.38)	$p = 0.056$
African Descent	145	0.94 (0.75-1.20)	
Mixed	407	1.37 (1.21-1.56)	
Gender			
Female	502	1.35 (1.20-1.53)	$\chi^2 = 7.764$
Male	424	1.05 (0.93-1.19)	$p = 0.005$
Age group			
18-30 years	481	2.08 (0.95-1.23)	$\chi^2 = 8.307$
31-40 years	271	1.27 (1.08-1.49)	$p = 0.016$
41-50 years	174	1.49 (1.24-1.79)	
Socio-economic position			
Low	578	1.11 (0.99-1.24)	$\chi^2 = 7.767$
Middle	307	1.45 (1.25-1.68)	$p = 0.021$
High	12	1.11 (0.59-2.10)	
Paternal/maternal history of diabetes			
None	757	1.22 (1.11-1.34)	$\chi^2 = 2.253$
Father, mother or both	113	1.22 (0.95-1.57)	$p = 0.954$
Paternal/maternal history of hypertension			
None	500	1.12 (0.99-1.26)	$\chi^2 = 7.633$
Father or mother	315	1.32 (1.14-1.52)	$p = 0.022$
Father and mother	55	1.79 (1.30-2.45)	
Paternal/maternal history of angina			
None	790	1.19 (1.08-1.30)	$\chi^2 = 4.097$
Father, mother or both	80	1.66 (1.22-2.25)	$p = 0.043$
Paternal/maternal history of MI			
None	824	1.20 (1.10-1.31)	$\chi^2 = 1.912$
Father, mother or both	46	1.73 (1.14-2.61)	$p = 0.167$
Paternal/maternal history of stroke			
None	824	1.22 (1.12-1.34)	$\chi^2 = 0.001$
Father, mother or both	46	1.20 (0.82-1.76)	$p = 0.988$

*Based on univariate analysis.

9.4.11 Allele frequencies of the tagging SNPS

Six SNPs in the CRP gene identified as tagging SNPs for African descent populations were genotyped and their allele frequencies were examined according to ancestral background. The –717A/G, -305A/G, +1444C/T and +2302G/A polymorphisms showed no deviation from the Hardy-Weinberg equilibrium (see Table 9.13). For the –286C/T/A polymorphism, there was significant deviation in the White-Hispanic group and the mixed group. However, there were no –286AA individuals in the White-Hispanic group, which may account for the high χ^2 values seen (p=0.001) and a larger sample size might reduce this distortion. In the mixed ancestry group, deviations from the Hardy-Weinberg equilibrium are more likely as the allele frequencies may be different. Similarly, this may also account for the deviation seen in this group for the +4899T/G polymorphism.

Substantial differences in genotype frequency were seen between the ancestral groups for all polymorphisms ($p<0.001$), except for the –717A/G SNP. The –305G/A SNP was not polymorphic in Amerindians, and the frequency of the A-allele ranged from 0.02 in White-Hispanics, to 0.16 in African descent subjects. The –286C/T/A polymorphism also exhibited large differences in the frequency of the A-allele, which was around 0.04 in White-Hispanics and Amerindians, and 0.35 in African descent individuals. The +1444T-allele had similar allele frequencies in White-Hispanic, Amerindian and mixed ancestry subjects of around 0.35, but was 0.16 in African descent subjects. In most populations, the +2302A-allele is the rare allele. In the ETNIAS data set, this allele had a frequency of 0.21 in African descent subjects, around 0.35 in the mixed ancestry and Amerindian subjects, and was 0.63 in White-Hispanics. The +4899G-allele also had large differences in allele frequency, with Amerindians having the lowest frequency of 0.02 and African descent subjects having the highest frequency of 0.35.

Qualitative analyses indicated that the allele frequencies were generally concordant with those from published data and available data from public domain resources. However, among the White-Hispanic group, for the +2302G/A polymorphism, the frequency of the G-allele (0.38) was lower and the frequency of the A-allele (0.63) was higher than that reported in other populations of European descent. This reversal of the MAF meant that the distribution was closer to that seen in Chinese and Japanese populations. These results provide strong evidence for differences in allele frequency between subjects of differing ancestral origin.

Table 9.13. Genotype frequency by ancestral background.

SNP	Ancestral background					χ^2 (p-value) across ancestral groups
	White-Hispanic	Amer-indian	African Descent	Mixed	Total	
-717A/G	n = 400	n = 63	n = 144	n = 421	N = 1028	
A-allele	0.799	0.810	0.865	0.827	0.820	6.67
G-allele	0.201	0.190	0.135	0.173	0.180	(0.083)
p-value for HW* deviation	0.042	1.000	0.724	1.000	0.169	
-305G/A	n = 396	n = 63	n =143	n =414	N = 1021	
G-allele	0.984	1.000	0.836	0.972	0.959	132.4
A-allele	0.016	-	0.164	0.028	0.041	(<0.001)
p-value for HW* deviation	0.095	1.000	0.767	1.000	0.080	
-286C/T/A	n = 396	n = 63	n = 143	n = 414	N = 908	
C-allele	0.578	0.632	0.441	0.541	0.545	110.3
T-allele	0.381	0.324	0.321	0.396	0.377	(<0.001)
A-allele	0.041	0.044	0.238	0.063	0.078	
p-value for HW* deviation	0.001	0.618	0.044	<0.001	<0.001	
+1444C/T	n = 400	n =63	n = 143	n = 416	N = 1027	
C-allele	0.653	0.651	0.836	0.659	0.681	36.84
T-allele	0.348	0.349	0.164	0.341	0.319	(<0.001)
p-value for HW* deviation	0.379	0.788	1.000	0.193	0.473	
+2302G/A	n = 400	n = 63	n = 143	n = 415	N = 1026	
G-allele	0.375	0.603	0.787	0.654	0.658	201.6
A-allele	0.625	0.397	0.213	0.346	0.341	(<0.001)
p-value for HW* deviation	0.202	0.601	0.619	0.067	0.489	
+4899T/G	n = 400	n = 63	n = 143	n = 416	N = 1027	
T-allele	0.954	0.976	0.650	0.917	0.898	229.9
G-allele	0.046	0.024	0.350	0.083	0.102	(<0.001)
p-value for HW* deviation	1.000	1.000	0.713	<0.001	<0.001	

* HW: Hardy-Weinberg equilibrium

9.4.12 Association of CRP genotypes with CRP concentration in different ancestral groups

All typed CRP polymorphisms were examined for association with CRP concentration. An initial analysis was conducted, which included all individuals regardless of ancestral background. No association was seen between the –717A/G polymorphism and CRP concentration (see Table 9.14). For the –305G/A variant, a difference of 6.91mg/L in CRP was seen between GG homozygous and AA homozygous individuals (p=0.007). This was the largest difference in CRP concentration seen in this cohort. Significant differences in CRP were also seen for the –286C/T/A and +1444C/T polymorphisms, although they were smaller in magnitude ($p \leq 0.025$). For the +2302G/A and +4899T/G SNPs, although the difference in CRP concentration by genotype did not reach statistical significance, the direction of the difference in CRP was consistent with that seen in prior studies, where individuals who were +2302AA homozygous had CRP concentrations 0.9mg/L lower than +2302GG homozygous individuals (see Chapters 6 and 7).

Individuals were then grouped by ancestral background and the association between genotype and CRP concentration was examined for each polymorphism (see Table 9.15). Due to the small size of the data set, no formal tests for interaction were performed, however, graphs were produced to allow a qualitative assessment of the potential for modification of effect size by ancestral group (see Figures 9.10-9.15).

For the –717A/G polymorphism, similar CRP concentrations were seen by genotype for the mixed ancestry group. For the White-Hispanic group, CRP concentrations appeared to be lower for the GG genotype compared to the AA genotype, with AG genotype subjects having intermediate concentrations. However, for the Amerindian and African descent subjects, no clear pattern was seen. Overall, there did not appear to be association with CRP concentration for any of the ancestral groups. For the –305G/A polymorphism, it was difficult to infer whether there were differences across the ancestral groups since only the White-Hispanic and African descent groups had individuals with all three genotypes. It appeared that AA subjects have higher CRP concentrations compared to GG subjects, but this difference was more extreme in the African descent group compared to the White-Hispanic group. Overall, there appeared to be association with CRP concentration across the ancestral groups in which this site was polymorphic, suggesting it may be functional.

When the –286C/T/A polymorphism was examined, the TT genotype appeared to have the highest CRP concentration in Amerindians, although it was not possible to examine the AA genotype due to lack of subjects. In the mixed ancestry group, AA subjects had the highest CRP concentrations, but in African descent subjects, the AA genotype had the lowest CRP concentrations and the TA genotype had the highest. Overall, the –286C/T/A polymorphism

showed broadly similar effects on CRP concentration across the ancestral groups, suggesting it may be functional For the +1444C/T polymorphism, the White-Hispanic and mixed ancestry groups showed a similar pattern of increase in CRP with CC subjects having the lowest concentrations and TT subjects having the highest and CT subjects having intermediate concentrations. TT genotype Amerindians also had the highest CRP concentrations, whereas African descent TT subjects had the lowest CRP concentrations. Overall, similar associations between the +1444C/T polymorphism and CRP concentration were seen across the ancestral groups except for the African descent group.

For the +2302G/A polymorphism, the mixed ancestry group had a clear pattern of decrease in CRP across the genotypic groups, with AA subjects having the lowest CRP concentrations. In the White-Hispanic group, similar concentrations were seen between the GA and AA genotypes, although they were lower than the GG genotype. For the Amerindian and African descent groups, this pattern appeared to be reversed, with the AA genotype having the highest CRP concentrations. Overall, similar associations were seen in the White-Hispanic and mixed ancestry groups between the +2302G/A SNP and CRP concentration. There appeared to be a reversal in the direction of association for the Amerindian and African descent groups. For the +4899T/G polymorphism, it was again difficult to infer whether there were differences across the ancestral groups since the White-Hispanic and Amerindian groups had no GG individuals. However, for the White-Hispanic group, the TG subjects had lower CRP concentrations compared to TT subjects, whereas Amerindian TG subjects had higher CRP concentrations. In the African descent group, the TG subjects had the highest CRP concentrations and GG subjects had the lowest. In the mixed ancestry group, small differences were seen between the TT and TG groups, and a similar CRP concentration was seen between the TG and GG groups.

Overall, differences in CRP concentration according to genotype were seen across the ancestral groups, and the pattern of increase or decrease in CRP concentration by genotype also varied by ancestral group.

Table 9.14. CRP concentrations according to CRP SNPs.

SNP	Number of Subjects	Geometric mean CRP (mg/L) (95% confidence interval)	p-value
-717A/G			
AA	634	1.27 (1.14-1.41)	0.322
AG	256	1.09 (0.92-1.29)	
GG	36	1.09 (0.74-1.61)	
-305G/A			
GG	844	2.79 (2.42-3.15)	0.007
GA	71	1.38 (1.02-1.87)	
AA	4	9.70 (2.12-44.38)	
-286C/T/A			
CC	201	0.91 (0.76-1.09)	0.004
CT	401	1.36 (0.20-1.55)	
CA	64	1.03 (0.70-1.51)	
TA	34	1.31 (0.83-2.09)	
TT	89	1.54 (1.20-1.99)	
AA	17	0.97 (0.40-2.35)	
+1444C/T			
CC	433	1.08 (0.95-1.23)	0.025
CT	393	1.29 (1.13-1.47)	
TT	99	1.53 (1.20-1.96)	
+2302G/A			
GG	408	1.33 (1.17-1.52)	0.101
GA	407	1.14 (1.00-1.31)	
AA	109	0.93 (0.32-0.64)	
+4899T/G			
TT	749	1.23 (1.12-1.35)	0.304
TG	151	1.16 (0.91-1.47)	
GG	25	0.92 (0.50-1.67)	

Table 9.15. CRP concentrations among CRP SNPs according to ancestral background.

SNP	Ancestral background											
	White-Hispanic			Amerindian			African Descent			Mixed		
	N	CRP	KW test*	N	CRP	KW test*	N	CRP	KW test*	N	CRP	KW test*
-717A/G												
AA	209	1.29 (1.08-1.54)	$\chi^2 = 1.616$	41	1.15 (0.70-1.88)	$\chi^2 = 2.129$	108	1.03 (0.79-1.35)	$\chi^2 = 1.609$	276	1.37 (1.17-1.61)	$\chi^2 = 0.030$
AG	89	1.08 (0.84-1.41)	p = 0.446	18	0.56 (0.26-1.25)	p = 0.344	33	0.69 (0.40-1.20)	p = 0.447	116	1.38 (1.09-1.75)	p = 0.985
GG	19	0.95 (0.53-1.70)		2	0.68		3	1.37 (0.31-2.58)		12	1.38 (0.82-2.34)	
-305G/A												
GG	305	1.18 (1.02-1.36)	$\chi^2 = 4.196$	61	0.92 (0.91-1.38)	N/A	99	0.82 (0.61-1.10)	$\chi^2 = 8.615$	379	1.36 (1.19-1.56)	$\chi^2 = 0.885$
GA	7	1.93 (0.67-5.55)	p = 0.123	0	-		41	1.16 (0.77-1.77)	p = 0.014	23	1.70 (1.01-2.85)	p = 0.348
AA	1	39.5		0	-		3	6.07 (3.38-10.9)		0	-	
-286C/T/A												
CC	66	0.79 (0.59-1.06)	$\chi^2 = 17.185$	11	1.12 (0.37-3.33)	$\chi^2 = 2.119$	20	0.65 (0.30-1.43)	$\chi^2 = 6.868$	104	1.04 (0.82-1.33)	$\chi^2 = 8.778$
CT	154	1.54 (1.29-1.88)	p = 0.002	16	0.75 (0.29-1.93)	p = 0.714	46	0.87 (0.60-1.26)	p = 0.231	185	1.44 (1.19-1.75)	p = 0.118
CA	20	0.60 (0.29-1.26)		2	1.31		25	0.99 (0.54-1.80)		17	1.98 (0.91-4.31)	
TA	4	1.46 (0.22-8.25)		1	1.52		13	1.92 (0.91-4.05)		16	1.24 (0.64-2.40)	
TT	30	1.29 (0.87-1.91)		2	3.33		11	1.48 (0.61-3.56)		46	1.69 (1.15-2.48)	
AA	0	-		0	-		11	0.65 (0.18-2.33)		6	2.02 (0.62-6.58)	
+1444C/T												
CC	124	0.99 (0.78-1.26)	$\chi^2 = 4.637$	25	0.94 (0.50-1.76)	$\chi^2 = 2.164$	100	0.93 (0.69-1.25)	$\chi^2 = 0.558$	184	1.27 (1.05-1.53)	$\chi^2 = 2.430$
CT	157	1.34 (1.10-1.63)	p = 0.098	29	0.77 (0.41-1.43)	p = 0.339	39	1.03 (0.64-1.64)	p = 0.756	168	1.43 (1.17-1.75)	p = 0.297
TT	36	1.53 (1.07-2.21)		7	1.72 (0.35 8.57)		4	0.63 (0.14-2.84)		52	1.61 (1.12-2.30)	
+2302G/A												
GG	123	1.37 (1.10-1.70)	$\chi^2 = 2.114$	22	0.95 (0.47-1.97)	$\chi^2 = 0.586$	87	0.95 (0.69-1.30)	$\chi^2 = 0.340$	176	1.61 (1.34-1.94)	$\chi^2 = 7.341$
GA	159	1.11 (0.91-1.37)	p = 0.348	28	0.80 (0.41-1.56)	p = 0.746	51	0.93 (0.63-1.37)	p = 0.844	169	1.32 (1.07-1.63)	p = 0.026
AA	35	1.13 (0.72-1.78)		11	1.16 (0.49-2.75)		5	1.07 (0.12-2.77)		58	0.97 (0.68-1.38)	
+4899T/G												
TT	285	1.26 (1.09-1.46)	$\chi^2 = 1.663$	58	0.90 (0.59-1.37)	$\chi^2 = 0.160$	59	0.87 (0.60-1.26)	$\chi^2 = 2.679$	347	1.34 (1.17-1.54)	$\chi^2 = 0.392$
TG	32	0.82 (0.46-1.47)	p = 0.197	3	1.38 (0.04-4.96)	p = 0.689	68	1.10 (0.77-1.56)	p = 0.262	48	1.56 (1.05-2.33)	p = 0.822
GG	0	-		0	-		16	0.67 (0.29-1.54)		9	1.61 (0.71-3.62)	

*KW test: Kruskall-Wallis test; N/A: not applicable

Figure 9.10. CRP concentration (95%CI) by –717A/G genotype and ancestral group.

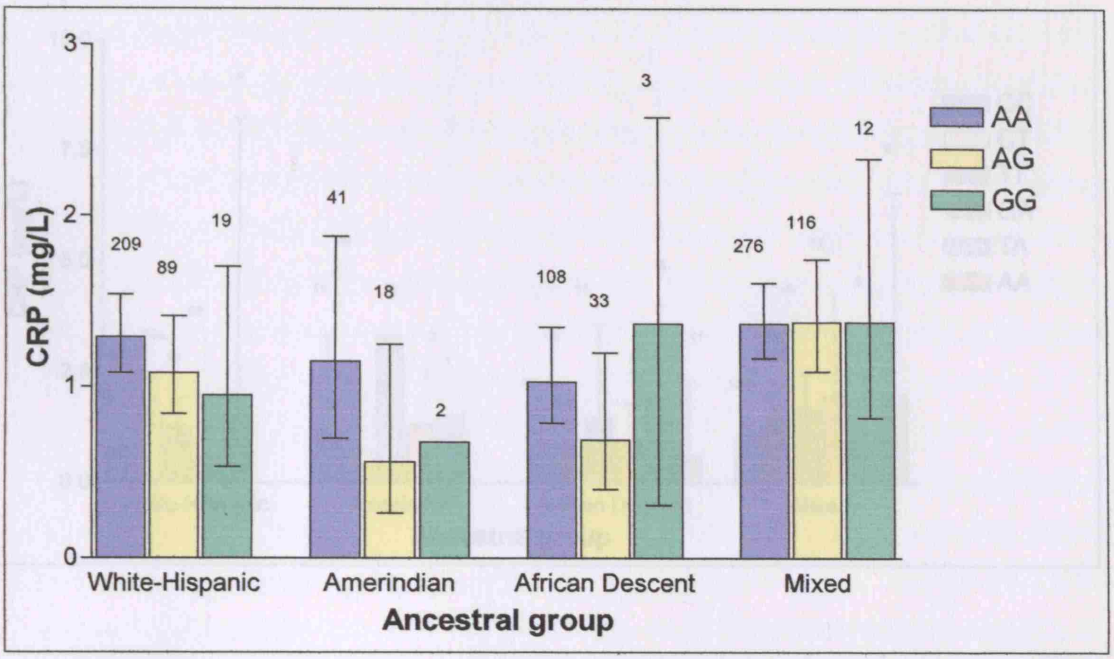


Figure 9.11. CRP concentration (95%CI) by –305G/A genotype and ancestral group.

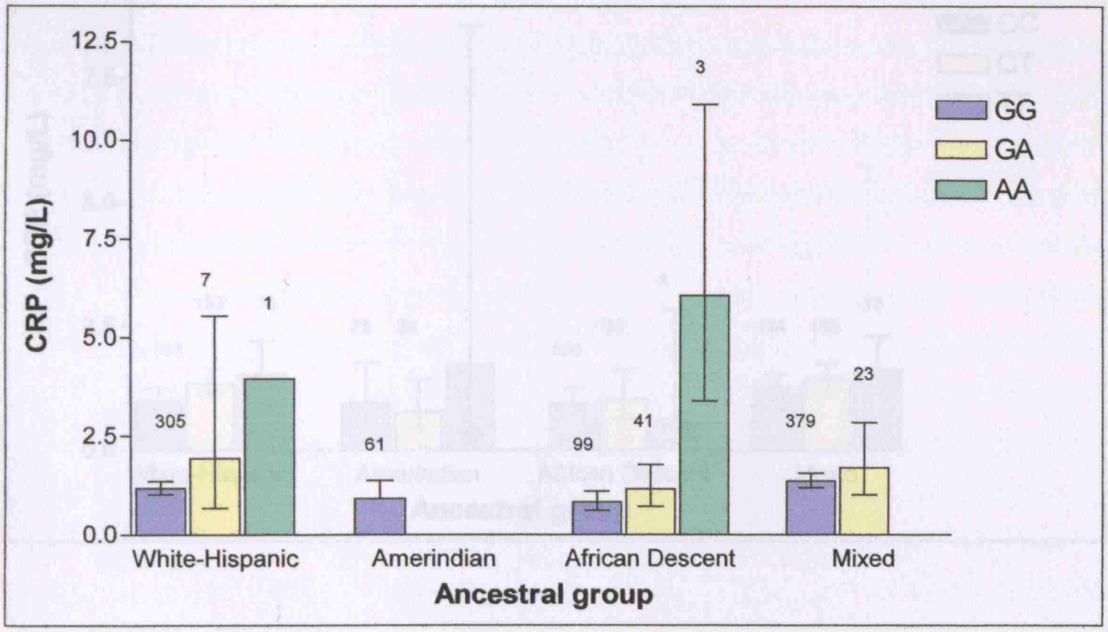


Figure 9.12. CRP concentration (95%CI) by –286C/T/A genotype and ancestral group.

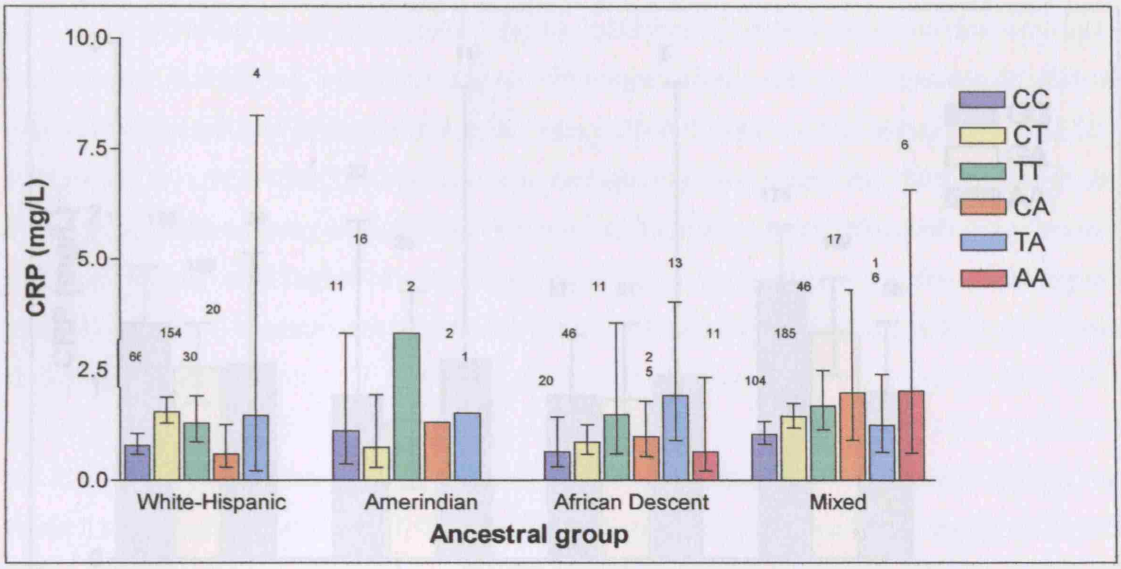


Figure 9.13. CRP concentration (95%CI) by +1444C/T genotype and ancestral group.

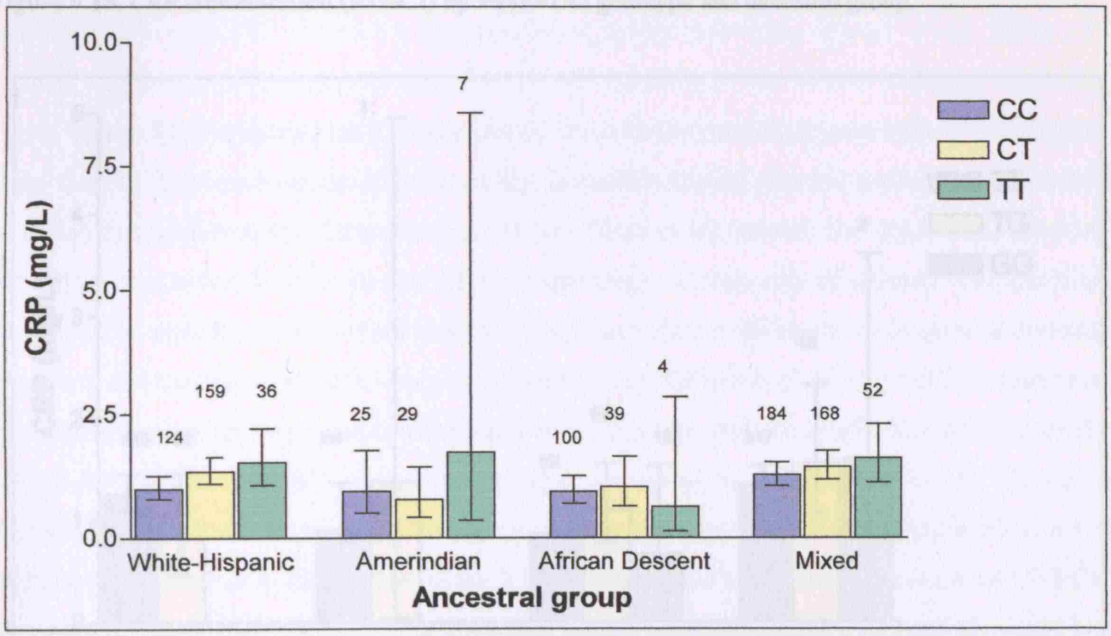


Figure 9.14. CRP concentration (95%CI) by +2302G/A genotype and ancestral group.

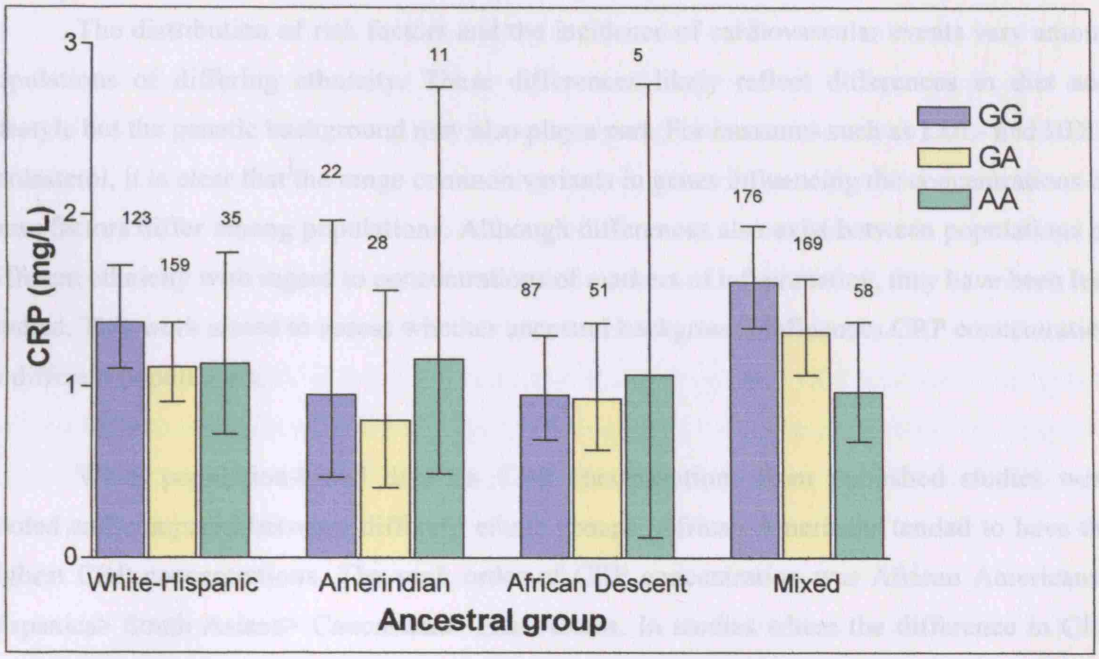
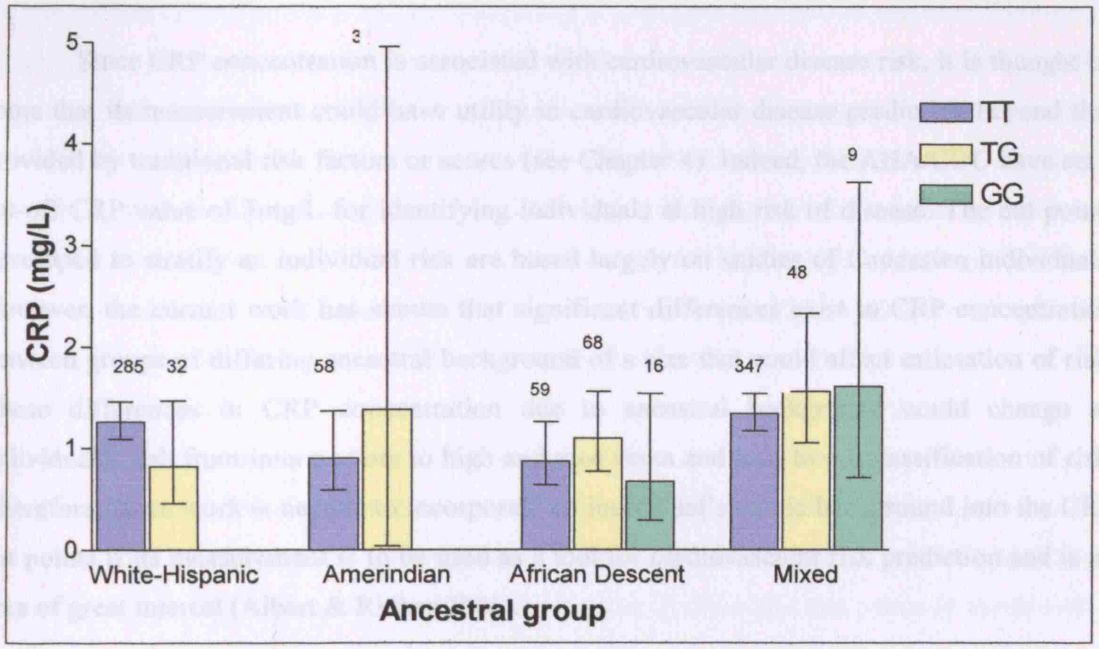


Figure 9.15. CRP concentration (95%CI) by +4899T/G genotype and ancestral group.



9.5 Discussion

The distribution of risk factors and the incidence of cardiovascular events vary among populations of differing ethnicity. These differences likely reflect differences in diet and lifestyle but the genetic background may also play a part. For measures such as LDL- and HDL-cholesterol, it is clear that the range common variants in genes influencing the concentrations of these factors differ among populations. Although differences also exist between populations of different ethnicity with regard to concentrations of markers of inflammation, they have been less studied. This work aimed to assess whether ancestral background influences CRP concentration in different populations.

When population-based data on CRP concentrations from published studies were pooled and compared between different ethnic groups, African Americans tended to have the highest CRP concentrations. The rank order of CRP concentration was African Americans> Hispanics> South Asians> Caucasians> East Asians. In studies where the difference in CRP concentration between a non-Caucasian group and Caucasian individuals was compared, non-Caucasians (African American, South Asian, East Asian and Hispanic) showed significant differences in mean CRP concentration. The direction of these differences identified in the comparator studies was consistent with the rank order derived from the population-based studies.

Since CRP concentration is associated with cardiovascular disease risk, it is thought by some that its measurement could have utility in cardiovascular disease prediction beyond that provided by traditional risk factors or scores (see Chapter 4). Indeed, the AHA/CDC have set a cut-off CRP value of 3mg/L for identifying individuals at high risk of disease. The cut points developed to stratify an individual risk are based largely on studies of Caucasian individuals. However, the current work has shown that significant differences exist in CRP concentration between groups of differing ancestral background of a size that could affect estimation of risk. These differences in CRP concentration due to ancestral background could change an individual's risk from intermediate to high and vice versa and lead to misclassification of risk. Therefore, more work is needed to incorporate an individual's ethnic background into the CRP cut points if its measurement is to be used as a tool for cardiovascular risk prediction and is an area of great interest (Albert & Ridker 2006).

The second part of this work was to determine whether the number and/or frequency of polymorphisms at the CRP gene locus, some of which are associated with CRP concentration, exhibit differences among groups of differing ancestry. Data obtained from both the SeattleSNPs and HapMap databases indicated a larger number of polymorphic sites among

subjects of African descent compared to individuals classified as Caucasian. In addition, of the SNPs common to more than one population, minor allele frequencies varied. For some SNPs, such as the +2302G/A variant (rs1205), there was a reversal of the common and rare alleles in the Chinese and Japanese samples compared to the European and African samples. Recent resequencing of the CRP gene to identify rare polymorphisms in additional populations also provided evidence for differences in the frequency of rare alleles at the CRP locus between ancestral groups. When the LD structure was examined, greater LD was seen between polymorphisms in the Caucasian subjects, with Mexican and Indo-Pakistani subjects having less LD, and African American subjects having the least LD between polymorphisms. This meant that Caucasian populations had the least number of haplotypes and African Americans had the greatest number of haplotypes. These results are consistent with data from HapMap, where the African descent genome is more highly polymorphic than Europeans, there is less LD between SNPs and thus more haplotypic groups.

Of particular interest in relation to questions about ancestral and ethnic differences in CRP concentration was the difference in frequency of the –286C/T/A promoter variant (rs3091244), which may be functional. An experimental study by Szalai et al. indicated that this SNP lies within an E-box binding site, which is created in the presence of the T-allele. In promoter-reporter gene assays, transcription of the luciferase reported in HepG2 cells was greater in the presence of the T- and A-alleles at the –286 site (Szalai *et al.* 2005). This polymorphism was identified in the SeattleSNPs database but not in the less SNP-dense HapMap. In European descent samples, the C-allele was the most common allele with a frequency of 0.62, followed by the T-allele (MAF=0.33) and then the rare A-allele (MAF=0.05). However, in African descent samples, the alleles were more evenly distributed with the T-allele being the most common allele with a frequency of 0.41, followed by the A-allele (MAF=0.33) and the C-allele is the rare allele (MAF=0.26). Since the T- and A-alleles appear to be associated with greater CRP promoter activity, this might account, in part, for the higher CRP concentrations seen in subjects of African descent.

From the resequencing work by Crawford et al., where Mexican and Indo-Pakistani population samples were also examined, only two alleles for the –286 variant were found in these populations, although this may be due to the small sample sizes for these populations (Crawford *et al.* 2006). Among Mexican samples, the C-allele was the common allele with a frequency of 0.75, and the T-allele was the minor allele. Among Indo-Pakistani samples, the C-allele was again the common allele, and the A-allele was the minor allele (MAF=0.17). Among the Southeast Asian samples in this study, all three alleles were seen. However, the allele frequency was very different to that seen in African Americans and Europeans, with the T-allele being the rarest allele, with a MAF of 0.06, and the A-allele being more common (MAF=0.19).

The most common C-allele had a frequency of 0.75, similar to that seen in the Mexican samples. Based on published studies on predominantly Caucasian subjects, individuals with the AA or TA genotype have the highest CRP concentrations (see Chapter 6).

Only three CRP variants and their association with CRP concentration have been studied in non-Caucasian populations. These are the (GT)n repeat variant in the CRP intron, the –286C/T/A polymorphism and the –305G/A polymorphism. The (GT)n repeat polymorphism has been studied in African Americans and Caucasians in the Carolina Lupus study (CLU) and in African Americans, Hispanics and Caucasians in the Lupus in Minorities: Nature vs. Nurture study (LUMINA) (Szalai *et al.* 2002; Szalai *et al.* 2005). Both of these studies have patients with systemic lupus erythematosus (SLE).

The (GT) bases can be repeated between 9 and 25 times within the intron, out of which the $(GT)_{16}$ and $(GT)_{21}$ are the most common in Caucasians. African American subjects had much lower frequencies of these alleles and differences in CRP concentration were also seen among these alleles (Szalai *et al.* 2002). In the LUMINA study, the $(GT)_{16}$ was also one of the most common alleles in all the ancestral groups studied. The $(GT)_{21}$ was the next common allele in Caucasians but not in African Americans or Hispanics. The $(GT)_{16}$ and $(GT)_{21}$ were associated with lower CRP concentrations compared to other alleles observed in the African American and Hispanic populations, resulting in differing CRP concentrations between the three ancestral groups (Szalai *et al.* 2005).

The –286C/T/A and –305G/A polymorphisms were also examined in the CLU study in both African American and Caucasian groups (Szalai *et al.* 2005). The –305G/A polymorphism is only present in African Americans with a minor allele frequency of 0.18 and was found to lie within the core consensus sequence for an E-box transcription factor binding site. Individuals with the common G-allele at this site have higher CRP concentrations compared to those with the A-allele. Since this site is not polymorphic in Caucasian subjects, the G base will always be present, resulting in the continuous presence of this binding site. Similarly, the T-allele of the –286C/T/A also results in the formation of an E-box, which may be weakly functional and individuals with the T-allele have higher CRP concentrations compared to individuals with the C-allele. This association was seen in both African American and Caucasian subjects, although the allele frequencies differed.

In order to examine the effect of the allele frequency differences seen from public domain resources and published work, on CRP concentration in subjects of differing ancestral background, a population-based study involving the genotyping of six CRP polymorphisms was conducted among Columbian subjects, selected on the basis of self-reported ancestry. These six

polymorphisms typed were chosen on the basis of their use as a tagging SNP set in an African descent population. Since, in general, subjects of African descent exhibit the greatest polymorphic variation, the lowest pairwise LD, and greatest haplotype diversity across the genome, these tagging SNPs should perform well in other populations.

Similar variation in minor allele frequency was seen for all polymorphisms among the White-Hispanic, native Amerindian, African Descent and mixed ancestral groups to that reported in the public domain resources and resequencing work, with one exception. The +2302G/A SNP (rs1205) had a low G-allele frequency of 0.375 and a high A-allele frequency of 0.625 in the White-Hispanic group, showing a reversal of the major and minor alleles of this polymorphism compared to subjects of European descent in the SeattleSNPs and HapMap databases. In addition, the –305G/A variant, which is not polymorphic in European populations, was seen in the White-Hispanic group with a MAF of 0.02 for the A-allele. This might indicate that the contemporary White-Hispanic group in Columbia may comprise an admixture of individuals of African and European descent and may not be a homogeneous group.

The association between CRP SNPs and CRP concentration was then examined. Broadly speaking, and with allowance for the small sample size, similar directional associations of CRP genotype and CRP concentration were observed overall. Individuals were next grouped according to ancestral background and the association between genotype and CRP concentration was re-evaluated. For the –717A/G SNP, no association with CRP concentration was seen in any ancestral group. For the –305G/A and -286C/T/A polymorphisms, similar effects on CRP concentration were seen across the ancestral groups. Since the frequency of these SNPs varies in different ancestral groups, and the association with CRP concentration appears to be retained across ancestral groups, these SNPs may be functional. Differences in CRP concentration were also seen according to the +1444C/T genotype; however, the effect was reversed in the African descent group. Similarly, associations were seen in the White-Hispanic and mixed ancestry groups between the +2302G/A SNP and CRP concentration and there appeared to be a reversal in the direction of association for the Amerindian and African descent groups. This may be because of small samples sizes or because these SNPs are a poor marker for the functional SNP(s) in the African descent group due to differences in the LD structure.

At the time of writing, haplotype analysis was in progress to assess the association of CRP haplotypes and CRP concentration across the ancestral groups. In addition, multivariate modelling to investigate the proportion of variance in CRP concentration accounted for by CRP genotype was being conducted. The aims of this study were also being extended to further data sets such as the Wandsworth Heart and Stroke Study that comprises 500 African descent individuals, 500 South Asian individuals and 500 European descent individuals.

Previous studies have shown that differences in allele frequency for common polymorphisms correlate with ancestral geographic location, and that self-identified ethnicity corresponds to the true ancestral origin in the majority of cases (Allocco *et al.* 2007; Tang *et al.* 2005; Lao *et al.* 2006). Broadly speaking, in the ETNIAS study, the minor allele frequencies for all the SNPs studied within each ancestral category were close to those expected from published data and public domain resources. Some of the genetic variants that differ in frequency between populations may be functional and this might account in part, for differences in phenotypes such as differences in levels of circulating blood factors, susceptibility to disease or response to treatments (Spielman *et al.* 2007). The work in this chapter found evidence of differences in allele frequency of CRP SNPs between populations of differing ancestry, in particular, differences in potentially functional SNPs. Since these SNPs are associated with CRP concentration, the differences in allele frequencies between groups of differing ancestry might reflect differences in selection pressure. Some SNPs may have been retained in African descent populations because they confer protection against infections that are more prevalent in that geographical location, but further work would be required to examine evidence of selection pressure in this region of the genome, and to ascertain if alleles encoding higher CRP concentrations do lead to protection from infectious disease.

9.6 Conclusions

Published data on CRP concentrations in different ancestral groups were examined and meta-analyses were conducted to compare CRP concentrations between Caucasian and other ancestral groups. Five ancestral populations were examined, Caucasian, African American, South Asian, East Asian and Hispanic. African American populations had a mean CRP concentration higher than Caucasian groups, and East Asian populations comprising Chinese and Japanese individuals, had lower CRP concentration compared to Caucasian populations. In subjects of differing ancestry in the ETNIAS data set comprising White-Hispanic, indigenous Amerindian, Afro-Caribbean and mixed ancestral groups from Columbia, CRP concentration also differed between groups of differing ancestry, that could reflect genetic, dietary, or other lifestyle differences between groups. When polymorphisms in and around the CRP gene were examined from public domain resources and resequencing work, greater polymorphic variation at this locus, lower pairwise SNP LD and a greater number of haplotypes were seen in African descent individuals compared to Caucasian, Chinese, Mexican and Indo-Pakistani populations. In addition, some SNPs were identified such as the –603A/G (rs3093061), -2C/T (rs3093064), -752G/A, +867C/T and +905C/T SNPs that appear to be unique to specific ancestral groups. Finally, minor allele frequencies of polymorphisms varied according to ancestral group. Similar variation in minor allele frequency was also seen in the ETNIAS data set. When the association

between CRP polymorphisms and CRP concentration was assessed, the –286C/T/A and +4899T/G variants had broadly similar associations across the ancestral groups, the +1444C/T variant had a decreased association with CRP concentration in African descent subjects, and the +2302G/A variant had a increased association with CRP concentration in Amerindian and African descent subjects. Where an allele shows similar association across populations of differing ancestry and varying LD, it is more likely to be functional. This work may have importance for the clinical utility of CRP as a predictive tool since CRP concentration appears to be influenced by ancestral background, and CRP polymorphisms that themselves affect CRP concentration are also modified by ancestry.

10. Use of Mendelian Randomisation to determine causality of CRP in coronary disease

10.1 Aim

To utilise a common single nucleotide polymorphism in the CRP gene that influences CRP concentration to assess whether the CRP-coronary heart disease association is likely to be causal.

10.2 Background

Recently, observational studies using high sensitive CRP assays have indicated that differences in plasma CRP concentrations within the normal range of 0.1-5mg/L are highly predictive of future coronary heart disease (CHD), an association that remains after adjustment for several risk factors. Individuals that fall in the highest tertile of normal CRP concentrations (a usual difference in CRP of about 2.4 vs. 1.0 mg/L), have an approximately 1.5 fold increase in risk of future vascular disease compared to individuals in the lowest tertile (Danesh *et al.* 2004). These observations have led to the proposal that CRP might play a causal role in atherosclerosis and its complications (Verma *et al.* 2005), a proposal supported by certain experimental studies of the effect of CRP on vascular cells and tissues, but these studies may be complicated by the effectd of contaminants in commercial CRP preparations (discussed in Chapter 1) (Scirica *et al.* 2006). Observational studies themselves do not provide unequivocal proof of causality, because an association between CRP and CHD may arise if CRP was mainly an inert marker for other causative exposures such as smoking status, blood pressure, diabetes, (with which it is correlated), or a marker of the inflammation that results from even subclinical atherosclerosis.

A randomised controlled clinical trial of a selective intervention to reduce CRP would provide an unbiased insight into the nature of the association. Although such a compound has recently been synthesised, it will be some time before even early stage clinical trials are possible (Pepys *et al.* 2006). Although inhibitors of HMG-CoA reductase (statins) are currently available are known to reduce CRP concentration, they also have major effects on cholesterol concentration and thereby give limited insight on the causal relevance of CRP per se (Balk *et al.* 2003).

A complementary and recently discussed approach to control for confounding and reverse causality bias is to utilise the natural randomisation of alleles at conception to minimise such biases. Such studies test associations of common polymorphisms in the CRP gene that are

reliably associated with differences in circulating CRP concentration. The inheritance of such variants should be subject to the random assortment of maternal and paternal alleles at the time of gamete formation, according to Mendel's second law, and this should balance the distribution of known and unknown confounders among genotypic classes. Since disease occurs after this process and thus has no effect on the randomised allocation of alleles, it also controls for reverse causality (Davey Smith & Ebrahim 2004; Morgan 1913). If CRP actually increases the risk of coronary events, then carriage of an allele that exposes individuals to a long-term elevation in CRP should confer an increased risk of coronary events proportional to the difference in CRP attributable to the allele (Davey Smith & Ebrahim 2004; Keavney 2002).

Several studies have already been conducted utilising this Mendelian randomisation approach to examine the potential causal relevance of circulating factors that have been implicated in CHD. For example, the association between homocysteine and CHD could be confounded in observational studies by factors such as smoking and blood pressure that are associated both with homocysteine level and CHD risk. However, polymorphisms exist in the methyltetrahydrofolate reductase (MTHFR) gene that are reliably associated with differences in serum homocysteine levels. In a meta-analysis, the +677TT genotype of MTHFR was associated with homocysteine concentrations higher by 1.7mmol/L than C-allele carriers, and with a higher risk of stroke compared to C-allele carriers (OR = 1.26), an effect that was concordant with that expected from the association of homocysteine with stroke risk in observational studies (Casas *et al.* 2005). The same may be true for homocysteine and coronary disease; however, a meta-analysis of 80 studies found that this association might be confounded by heterogeneity due to geographical variation and publication bias might be present, although this is still an area of debate (Lewis *et al.* 2005; Wald *et al.* 2006).

Mendelian randomisation has also been used to examine the relationship between fibrinogen and CHD. Case-control and prospective studies showed that a 0.14g/L increase in fibrinogen was associated with a relative risk of 1.17 for CHD. A polymorphism in the promoter region of the β-fibrinogen gene was shown to alter plasma fibrinogen levels by around 0.14g/L. Therefore, if fibrinogen was a causal factor in CHD, the polymorphism should be associated with a relative risk for CHD of around 1.17. However, in a meta-analysis of these studies, the risk of CHD associated with this genotype was only 1.0, suggesting observational studies of fibrinogen and CHD may also be affected by confounding (Keavney *et al.* 2006). However, one limitation of the fibrinogen analysis is that in contrast to the case with CRP, amino acid variants of fibrinogen exist so that β-fibrinogen gene variants that only influence fibrinogen level may not adequately capture their functional variation.

The +1444C/T polymorphism (rs1130864) in the 3' untranslated region (UTR) of the CRP gene has already been shown to be robustly associated with higher basal and stimulated CRP concentrations in several published studies and studies described in Chapters 6, 7 and 8, with no effect on a range of other covariables relevant to CHD. This SNP was therefore chosen for the Mendelian randomisation approach to determine causality. Although any of the SNPs in the CRP gene with effects on CRP concentration (or haplotypes inferred from typing tagging SNPs) could have been utilised, the +1444C/T variant was chosen, as this was the most studied SNP at the time the Mendelian randomisation analyses were planned.

10.3 Methods

10.3.1 Study Populations

Genotyping for the CRP polymorphism and measurement of plasma CRP concentration were conducted in samples obtained from a number of cross-sectional or prospective studies, or randomised controlled trials as described in Table 10.1 (see also Chapter 3). To test the association of the +1444C/T genotype with CRP concentration, male subjects without clinically evident cardiovascular disease from six studies with available DNA and plasma samples were studied. These were the NPHSII, WOSCOPS, HIFMECH, Army, UDACS and EBCT studies. A second analysis was conducted to evaluate the association between genotype and risk of MI in male subjects from four studies (NPHSII, LEADER, WOSCOPS and HIFMECH). Cohorts are described in brief below.

NPHSII cohort

The NPHSII cohort is a large prospective study of 3,012 healthy Caucasian men originally recruited in 1986. Nine general practices participated in the study, and all patients were free of a history of unstable angina, MI or evidence of silent infarction, coronary surgery, aspirin or anticoagulant therapy, cardiovascular disease, or malignant disease at the time of recruitment. Baseline characteristics and demographic information were ascertained by means of a clinical assessment and a questionnaire completed at the beginning of the study. The endpoints were fatal CHD events and non-fatal MI, coronary artery surgery and silent MI on follow-up ECG. At present, 227 CHD events have been recorded. Genotyping was carried out on 2,676 men; the rest were non-Europeans or had incomplete measurements and were therefore ineligible.

WOSCOPS cohort

The West of Scotland Coronary Prevention Study consists of 6,595 moderately hypercholesterolemic men recruited by population screening. The principal exclusion criteria were previous MI, angina pectoris requiring hospitalisation within the previous 12 months, and

life-threatening non-cardiac illness. 580 individuals during the course of the follow-up experienced a fatal or non-fatal MI, sudden coronary death or required coronary artery bypass graft or angioplasty and were defined as cases. They were matched with 2 controls drawn from the cohort on the basis of age and smoking status. Participants were randomised to either receive 40 mg Pravastatin each evening or a placebo. They were followed up every 3 months for a mean of 4.9 years. End points were identified from information received at routine trial visits, by analysis of annual electrocardiograms and from death reports. Genotyping was carried out by the WOSCOPS group on 1,596 men from the placebo group, from which 348 had reported an event.

HIFMECH cohort

The Hypercoagulability and Impaired Fibrinolytic function MECHanisms predisposing to MI (HIFMECH) study is a large European multicentre case-control study, comprising Caucasian, male survivors of a MI, and age-matched controls from the same regional areas as cases. Men were recruited from four centres: Stockholm, London, Marseilles and San Giovanni Rotondo. Patients were recalled for assessment between 3 and 6 months after the MI. At the time of the present study, 491 patients and 517 controls were genotyped in the respective centres.

Army volunteer exercise study

This cohort comprised 250 healthy Caucasian individuals from the Army Training Regiment, Bassingbourn, UK who were initially recruited at the beginning of an 11-week period of basic training. The end of the training was marked with an intensive 48-hour final military endurance exercise (FME). Blood samples were then taken on three occasions (2, 48 and 96 hours) following return from the FME. DNA had been previously extracted from all individuals, genotyped and analysed (Brull *et al.* 2003).

Electron Beam Computerised Tomography (EBCT) Study in Diabetics

A random sample of Type 1 diabetic individuals aged 30 to 55 years was taken from the diabetes registers of five London hospitals. A random sample of the general population (94 men and 107 women), stratified to have a similar age and gender distribution to the patients with diabetes, was drawn from the lists of two London general practices (controls). A questionnaire was completed and height, weight, waist:hip ratio and blood pressure were measured using standard methods. EBCT was used to compare coronary artery calcification and coronary risk factors in Type 1 diabetic patients and non-diabetic participants. Participants were followed up one year after the initial testing. For the present study only 74 male controls with CRP levels were included and their genotypes were available for analysis.

LEADER cohort

The Lower Extremity Arterial Disease Event Reduction (LEADER) trial is a double-blind randomised controlled trial, carried out in men with lower extremity arterial disease (LEAD). 1,568 individuals were recruited to investigate the effects of Bezafibrate in the prevention of CAD events. Individuals with a previous MI or stroke were eligible for the trial provided that their general management was stable. Active treatment consisted of Bezafibrate 400 mg/day for men with creatinine levels below 135 μmol/l and was placebo controlled and double-blind. Participants were seen 1 month after treatment started and then at 3 month intervals. Blood was taken for measurements twice at baseline, once at 1 and 3 months and then once at six monthly intervals. Genotyping and analysis was carried out on 647 Caucasian men in active treatment and 419 on placebo.

UDACS cohort

The UCL Diabetes and Cardiovascular disease Study (UDACS) study is a cross sectional case-control study aimed to evaluate risk factors for CVD on 1,020 subjects with diabetes. Patients were recruited from the diabetes clinic at University College London Hospitals NHS Trust (UCLH) between the years 2001–2. Analyses were confined to Caucasian subjects with Type 2 diabetes. CVD was recorded if a patient had one or more of CHD, peripheral vascular disease, or cerebrovascular disease. For the current study 348 Caucasian males without coronary heart disease were included.

Table 10.1. Design characteristics of the original studies included in the present study.

Study name	Study Design & Median follow-up (years)	Sample size Original/ Genotyped	Country	Study population†	Main exclusion crite
NPHSII	Prospective Cohort, (10.6)	3012/2676	UK*	Healthy middle-aged men.	Pre-existing cardiovascula disease, Coronary surgery, Aspirin or anticoagulant therapy, Malignant disease.
LEADER	Nested case-control study from a clinical trial of bezafibrate treatment, (4.6)	1568/1066	UK	Men with lower extremity arterial disease.	Unstable angina, Total cholesterol <3.5 or > mmol/l., Significant renal or hepati disease or malignant disea
WOSCOPS	Nested case-control in a clinical trial of pravastatin treatment, (4.9)	6595/1596	UK	Moderately hypercholesterolemic men.	MI, or angina pectoris req hospitalisation, Life-threatening non-cardi illness.
HIFMECH	Case-control, (N/A)	Cases: 533/491 Controls: 575/517	UK, Sweden, France, Italy	Cases: male MI survivors. Controls: men matched by age and regional areas.	Familial hypercholesterola Insulin-dependent DM‡.
Army	Cross-sectional, (N/A)	250/227	UK	Healthy UK-Army recruits.	N/A
UDACS	Case-control, (N/A)	Controls: 449/348	UK	Cases: men and women with DM‡ and cardiovascular disease. Controls: men and women with DM but without cardiovascular disease.	N/A
EBCT	Case-control (N/A)	Controls: 94/74	UK	Cases: men and women with type-1 DM. Controls: healthy men and women matched by age and gender.	N/A

*United Kingdom. †All subjects evaluated were Caucasians. ‡DM: diabetes mellitus.

10.3.2 Data Collection

Data on sex, mean age, systolic and diastolic blood pressure, body mass index, smoking status, glucose, lipid profile, alcohol consumption, fibrinogen and plasma CRP values were obtained from the original studies. Subjects were classified using unified definitions of hypertension, hypercholesterolaemia, type-II diabetes mellitus and obesity from the guidelines on primary prevention of the American Heart Association (Pearson *et al.* 2002). For the studies relating genotype and coronary events, non-fatal MI according to WHO criteria was considered the primary outcome, as this end-point had been uniformly used across all studies (Tunstall-Pedoe *et al.* 1994).

10.3.3 Genotyping of the +1444C/T CRP polymorphism

Genotyping of the +1444C/T polymorphism was conducted by PCR-RFLP analysis (see Chapter 3) using the restriction endonuclease *SduI* that recognises the common allele, but not the rare allele. The forward primer sequence was mismatched to force an allele-specific restriction enzyme site into the PCR product. The region containing the relevant polymorphism was amplified by PCR. Digestion products were electrophoresed at 100V for 1 hour, and visualised by ethidium bromide fluorescence under U.V. light. Subsequently, 200 samples were re genotyped in the NPHSII cohort using a standard TaqMan assay to establish consistency between the two methods.

10.3.4 Measurement of CRP

CRP concentrations for the NPHSII (717 subjects) and LEADER studies were measured by use of commercial assays (R&D Systems). Inter-assay and intra-assay coefficients of variations were 6.2% and 1.9% respectively, with a detection limit of 0.1 mg/L. For an additional 2221 subjects from the NPHSII study, an enzyme immunoassay (Kordia Life Sciences) was used to measured CRP concentrations. For the EBCT and HIFMECH studies CRP was measured with a highly sensitive in-house enzyme-linked immunosorbent assay with rabbit anti-human CRP (Dako, Copenhagen, Denmark) as a catching and tagging antibody with inter-assay and intra-assay coefficients of variations of 4.7 and 3.8% and a limit of detection of 0.15mg/L. For the WOSCOPS study a validated in-house assay was used with a lower limit of detection of 0.1 mg/L. The inter-assay and intra-assay coefficients of variation were 6.2% and 1.9% respectively. In the Army-study, CRP was measured on a BN Prospec (Dade Behring, Milton Keynes, UK). Inter-assay and intra-assay coefficients of variation were <4% and <2% respectively with a detection limit of 0.2 mg/L. In the UDACS study, CRP was measured using a highly sensitive ELISA assay (Dako A/S, Glostrup, Denmark). Inter-assay and intra-assay coefficients of variation were 8% and 10% respectively with a detection limit of 0.26 mg/L. All CRP values had been previously measured and were available from each study for analysis.

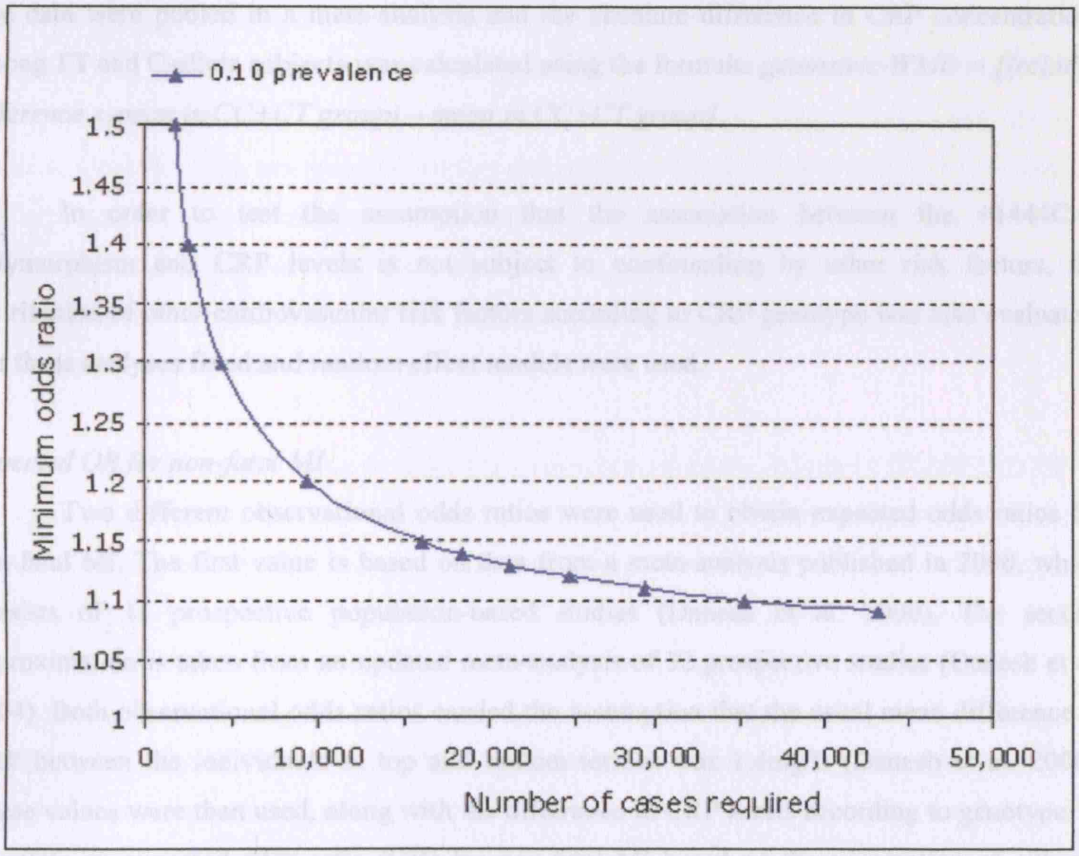
10.3.5 Statistical methods and analyses

These analyses were planned and conducted by Dr. Juan Pablo Casas and Dr. Leonelo Bautista (University of Wisconsin). I contributed to the preparation of tabular data and helped to conduct some of the analyses under Dr. Casas' supervision.

Sample size calculation

A sample size calculation was conducted to determine the minimum odds ratio that could be reliable detected for a given number of cases and controls in order to achieve robust Mendelian randomisation analyses of the +1444C/T genotypes at the 0.01 significance level and with 90% power (see Figure 10.1). The prevalence of the +1444TT genotype is around 0.10 (MAF=0.3) in European populations. Since it was not possible to obtain the large number of cases, all available studies were assessed.

Figure 10.1. Sample size calculation to determine the number of cases required for Mendelian randomisation analyses (produced by Professor Liam Smeeth, London School of Hygiene and Tropical Medicine).



This sample size calculation involves the following assumptions: (i) the impact of CRP genotype on CRP levels is about 0.7mg/L (derived from published studies) and (ii) the relative risk for CHD is about 1.5 per 1.4mg/L increase in measured CRP levels (derived from the

largest single study thus far and a literature-based meta-analysis of 22 prospective studies (Danesh *et al.* 2004)).

Genotype and CRP concentration

To quantify the effect of variation in the CRP gene on CRP concentration, 4659 men from six studies (NPHSII, WOSCOPS, HIFMECH, Army, EBCT and UDACS) were genotyped. This analysis was limited to individuals without known coronary or peripheral artery disease in order to avoid the potential modifying effect of disease on the genotype-CRP association. For prospective studies, baseline CRP data were used from all available subjects without clinically evident atherosclerosis. For case-control studies, genotype-CRP associations were analysed solely using control groups. Because of the skewed distribution of the CRP values, logarithmic means and their standard deviations were obtained.

The within-study mean difference in CRP concentration was calculated between individuals homozygous for the T-allele and carriers of the C-allele and then each mean was weighted by the inverse of its variance to obtain an overall weighted mean difference (WMD). The data were pooled in a meta-analysis and the absolute difference in CRP concentrations among TT and C-allele subjects was calculated using the formula: *geometric-WMD = [(relative difference x mean in CC+CT group) – mean in CC+CT group].*

In order to test the assumption that the association between the +1444C>T polymorphism and CRP levels is not subject to confounding by other risk factors, the distribution of other cardiovascular risk factors according to CRP genotype was also evaluated. For these analyses fixed and random effect models were used.

Expected OR for non-fatal MI

Two different observational odds ratios were used to obtain expected odds ratios for non-fatal MI. The first value is based on data from a meta-analysis published in 2000, which consists of 11 prospective population-based studies (Danesh et al. 2000). The second approximation is taken from an updated meta-analysis of 22 prospective studies (Danesh et al. 2004). Both observational odds ratios carried the assumption that the usual mean difference in CRP between the individuals in top and bottom tertiles was 1.4mg/L (Danesh et al. 2000). These values were then used, along with the difference in CRP levels according to genotype, to calculate an expected odds ratio (OR) for non-fatal MI based on the assumption of either a linear or exponential relationship between CRP and non-fatal MI.

The expected OR for TT homozygous subjects compared to C allele carriers was calculated using the formula:

Expected OR= OR-non genetic studies $^{(WMD/1.4)}$ where WMD = weighted mean difference between TT subjects and C-allele carriers

The 95% confidence interval for this expected OR was obtained by simulation. One million replications of the expected OR were obtained using the WMD in CRP by genotype and the OR from the non-genetic observational studies with the corresponding standard errors. The values for the 2.5 and 97.5 percentiles of the simulated distribution were used as limits for the 95% confidence interval for the expected OR. This expected OR was compared with the summary-OR obtained from the genetic studies by means of an interaction test (Altman & Bland 2003), and any consistency would suggest that the association between CRP and coronary events identified in prior observational studies would be unlikely to be the result of residual confounding or reverse causality bias.

Genotype and non-fatal MI

To examine the effect of CRP genotype on risk of non-fatal MI, 6201 men from four studies (NPHSII, WOSCOPS, HIFMECH and LEADER) were genotyped, and the individual adjusted-ORs and 95% confidence interval (CI) for subjects homozygous for the T allele compared with carriers of the C allele were obtained. The variables included in the multivariate model for each study were age, hypertension, total-cholesterol, body mass index, diabetes mellitus, alcohol intake and smoking status. Although the distribution of potential confounders was expected to be balanced among the genotypic groups, adjusted analyses were undertaken to narrow the confidence limits of the effect size. In addition, for the two intervention trials (WOSCOPS and LEADER), potential interaction of the genotype-MI association with the active therapy used in each trial was also evaluated.

These four studies were then pooled to obtain a summary adjusted-OR and 95% CI for non-fatal MI, under both fixed and random effect models. Fixed effect summary-ORs were calculated using the inverse variance-weighted method (Shadish & Haddock 1994) and the DerSimonian and Laird method (DerSimonian & Laird 1986) was used to calculate the random effect summary OR. The DerSimonian and Laird Q test was used to evaluate the degree of heterogeneity between studies and I^2 was used as a measure to describe the percentage of variability in point estimates due to heterogeneity rather than sampling error (Higgins *et al.* 2003).

Consistency between expected and observed ORs

Consistency between expected and observed ORs for non-fatal MI was examined using an interaction test (Altman & Bland 2003). Consistency between these ORs would suggest that the association between CRP and non-fatal MI identified in prior observational epidemiological

studies is unlikely to be the result of confounding or reverse causality bias. A sensitivity analysis was also conducted to evaluate the robustness of the main comparison, based on several assumptions, such as genetic model of inheritance (co-dominant), study design (prospective studies included only), and different estimates of the mean increase in the risk (lower: 60% and higher: 150%) of coronary events reported from the previous meta-analysis of prospective studies.

Data were analysed using the Review Manager software (version 4.2) from the Cochrane Collaboration 2003 and Stata 8.2 (Stata Corporation, College Station, Texas, 2003).

10.4 Results

10.4.1 +1444C/T genotypes and allele frequencies

The allele and genotype frequencies of the +1444C/T polymorphism for all the studies were in Hardy-Weinberg equilibrium, except for a very small distortion seen in the control group in the NPHSII study. The frequencies of the rare allele in disease-free subjects from all studies were very similar (range: 26% to 33%) (see Table 10.2).

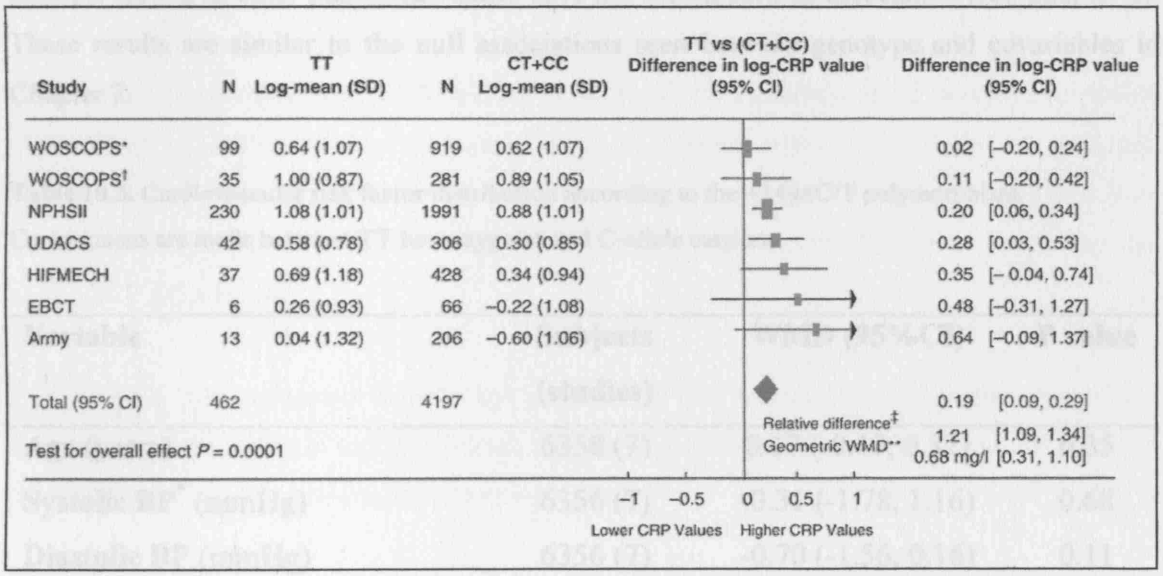
Table 10.2. CRP +1444C/T genotype and allele frequency.

Study	Genotype	Controls (number)	Cases (number)
NPHSII			
Prospective, observational	+1444CC	1261	57
Nested case - control	+1444CT	1047	41
	+1444TT	265	5
	T-allele frequency	0.30	0.24
	H-W p value	0.03	0.48
WOSCOPS			
Prospective, observational	+1444CC	511	175
Nested case - control	+1444CT	481	136
(substudy from RCT)	+1444TT	111	37
	T-allele frequency	0.32	0.30
	H-W p value	0.88	0.17
LEADER			
Prospective, observational	+1444CC	514	21
Nested case - control	+1444CT	412	16
(substudy from RCT)	+1444TT	97	6
	T-allele frequency	0.29	0.32
	H-W p value	0.27	0.31
HIFMECH			
Case - control	+1444CC	254	246
	+1444CT	224	196
	+1444TT	39	49
	T-allele frequency	0.29	0.30
	H-W p value		
ARMY			
Cross - sectional	+1444CC	122	-
	+1444CT	92	-
	+1444TT	13	-
	T-allele frequency	0.26	-
	H-W p value	0.42	
UDACS			
Case - control	+1444CC	167	-
	+1444CT	139	-
	+1444TT	42	-
	T-allele frequency	0.32	-
	H-W p value	0.12	
EBCT			
Cross - sectional	+1444CC	39	-
	+1444CT	29	-
	+1444TT	6	-
	T-allele frequency	0.28	-
	H-W p value	0.85	

10.4.2 Relationship between CRP genotype and basal concentration

Six studies were analysed to examine the relationship between CRP genotype and basal concentration. This involved data from 4659 subjects from WOSCOPS, NPHSII, HIFMECH, UDACS, EBCT and the Army studies. The weighted mean CRP concentration in C-allele carriers without known cardiovascular disease was 2.01 mg/L (95%CI: 1.94-2.07). Using a fixed effects model, the geometric-weighted mean difference (WMD) in plasma CRP concentration between men homozygous for the T-allele compared with C-allele carriers was calculated to be 0.68 mg/L (95%CI: 0.31-1.10, p=0.0001) (see Figure 10.2). No significant heterogeneity was seen between the studies (p=0.47). These results are concordant with those obtained from other published studies (see Chapter 6) and other unpublished studies (see Chapter 7) examining associations of the +1444C/T SNP with CRP concentration.

Figure 10.2. Relative difference and geometric-WMD in plasma CRP according to CRP genotype.



*Data for Control subjects.

†Data for subjects with non-fatal MI.

‡Relative difference was obtained by antilog of the difference in log-CRP values.

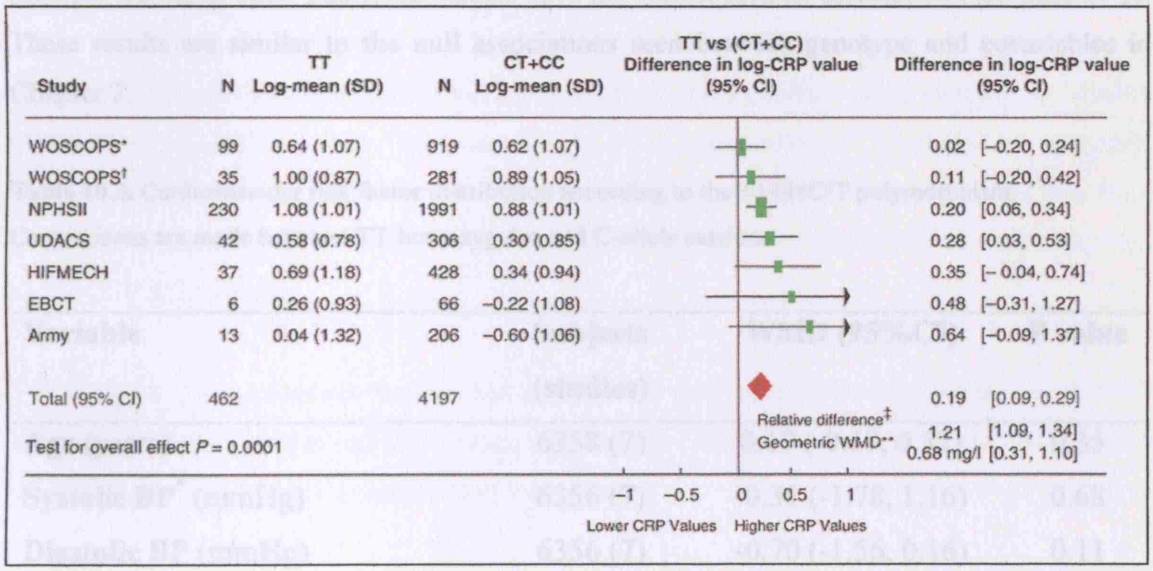
**Geometric WMD: weighted mean difference.

When alternative models of the effect of genotype on CRP were evaluated, only the (TT) vs. (CC) comparison was significant (WMD=0.63 mg/L, 95%CI: 0.25-1.06, p=0.0007), while heterozygosity (CT vs. CC) was not (WMD=0.03, 95%CI: -0.18-0.22, p=0.85). Although some individual studies had a clear co-dominant pattern, overall, the pooled data was consistent with

10.4.2 Relationship between CRP genotype and basal concentration

Six studies were analysed to examine the relationship between CRP genotype and basal concentration. This involved data from 4659 subjects from WOSCOPS, NPHSII, HIFMECH, UDACS, EBCT and the Army studies. The weighted mean CRP concentration in C-allele carriers without known cardiovascular disease was 2.01 mg/L (95%CI: 1.94-2.07). Using a fixed effects model, the geometric-weighted mean difference (WMD) in plasma CRP concentration between men homozygous for the T-allele compared with C-allele carriers was calculated to be 0.68 mg/L (95%CI: 0.31-1.10, p=0.0001) (see Figure 10.2). No significant heterogeneity was seen between the studies (p=0.47). These results are concordant with those obtained from other published studies (see Chapter 6) and other unpublished studies (see Chapter 7) examining associations of the +1444C/T SNP with CRP concentration.

Figure 10.2. Relative difference and geometric-WMD in plasma CRP according to CRP genotype.



*Data for Control subjects.

†Data for subjects with non-fatal MI.

‡Relative difference was obtained by antilog of the difference in log-CRP values.

**Geometric WMD: weighted mean difference.

When alternative models of the effect of genotype on CRP were evaluated, only the (TT) vs. (CC) comparison was significant (WMD=0.63 mg/L, 95%CI: 0.25-1.06, p=0.0007), while heterozygosity (CT vs. CC) was not (WMD=0.03, 95%CI: -0.18-0.22, p=0.85). Although some individual studies had a clear co-dominant pattern, overall, the pooled data was consistent with

a recessive pattern and therefore, a recessive model of the effect of genotype on CRP concentration was inferred for these analyses.

10.4.3 Genotype, CRP concentration and confounding

In order to test the assumption that the association between the +1444C/T polymorphism and CRP levels is not subject to confounding by other risk factors, which is a prerequisite of the Mendelian randomisation approach, the distribution of other cardiovascular risk factors according to CRP genotype was also evaluated. These risk factors included age, systolic and diastolic blood pressure, body mass index, current smoking, glucose, alcohol intake, fibrinogen, triglyceride, total and HDL cholesterol. Using a recessive model of inheritance, no significant difference was seen for any of these variables (see Table 10.3), except for BMI, which was slightly higher among TT homozygous individuals by 0.34 kg/m^2 (95%CI: 0.06-0.62, p=0.02). However, this may be a positive result as a consequence of multiple testing as other published studies have not shown such an association (see Chapter 6). These results are similar to the null associations seen between genotype and covariables in Chapter 7.

Table 10.3. Cardiovascular risk factor distribution according to the +1444C/T polymorphism. Comparisons are made between TT homozygotes and C-allele carriers.

Variable	Subjects (studies)	WMD (95%CI)	P value
Age (years)	6358 (7)	0.17 (-0.19, 0.52)	0.35
Systolic BP* (mmHg)	6356 (7)	-0.31 (-1.78, 1.16)	0.68
Diastolic BP (mmHg)	6356 (7)	-0.70 (-1.56, 0.16)	0.11
Body mass index (kg/m^2)	6359 (7)	0.34 (0.06, 0.62)	0.02
Total cholesterol (mmol/L)	6115 (6)	-0.01 (-0.08, 0.06)	0.87
HDL-cholesterol (mmol/L)	4714 (5)	-0.01 (-0.04, 0.01)	0.32
Triglyceride (mmol/L)	6041 (5)	-0.03 (-0.11, 0.06)	0.51
Fibrinogen (g/L)	5656 (4)	0.03 (-0.02, 0.08)	0.20
Glucose (mmol/L)	1913 (3)	-0.01 (-0.13, 0.12)	0.90
Alcohol intake (U/week)	4549 (4)	0.01 (-1.38, 1.39)	0.99
Current smoking†	6132 (6)	1.00 (0.84, 1.20)	0.98
C-reactive protein (mg/L)	4659 (6)	0.68 (0.31, 1.10)	0.0001

*BP: Blood pressure.

†Instead of the WMD, the value reported is the odds ratio.

10.4.4 Risk estimates from observational studies

In order to estimate the expected OR for non-fatal MI, data from prior non-genetic observational studies that have examined the CRP-coronary event association have been used. Over time, as more studies have been carried out, the strength of this association has been reduced, as larger data sets have been used thus reducing random error, and better adjustment has been used, therefore two values for the expected OR were estimated so as to compare the differences in the data over time.

The first value is based on data from a meta-analysis published in 2000, which consists of 11 prospective population-based studies (Danesh *et al.* 2000a) involving 1953 cases, from mainly Caucasian men. The weighted mean follow-up is 8 years and the studies have been adjusted for traditional cardiovascular risk factors and regression dilution bias. This study reported a RR for coronary events of 2.0 (95%CI: 1.6-2.5) for individuals in the top vs. bottom tertiles of the CRP distribution.

The second estimate is taken from an updated meta-analysis of 22 prospective studies (Danesh *et al.* 2004), with 7068 cases that have been adjusted similarly for potential confounders of cardiovascular disease but have not been adjusted for regression dilution bias. This meta-analysis gives an OR of 1.58 (95%CI: 1.48-1.68), also for the top vs. bottom tertiles of the CRP distribution.

Both these estimates carried the assumption that the usual mean difference in CRP between the individuals in top and bottom tertiles was 1.4mg/L (Danesh *et al.* 2000a), and that the CRP-coronary event relationship was log-linear.

Table 10.4. Expected odds ratios for non-fatal MI carriers based on a between-genotype WMD difference in CRP of 0.68 mg/L (95%CI: 0.31-1.10) calculated from observational studies.

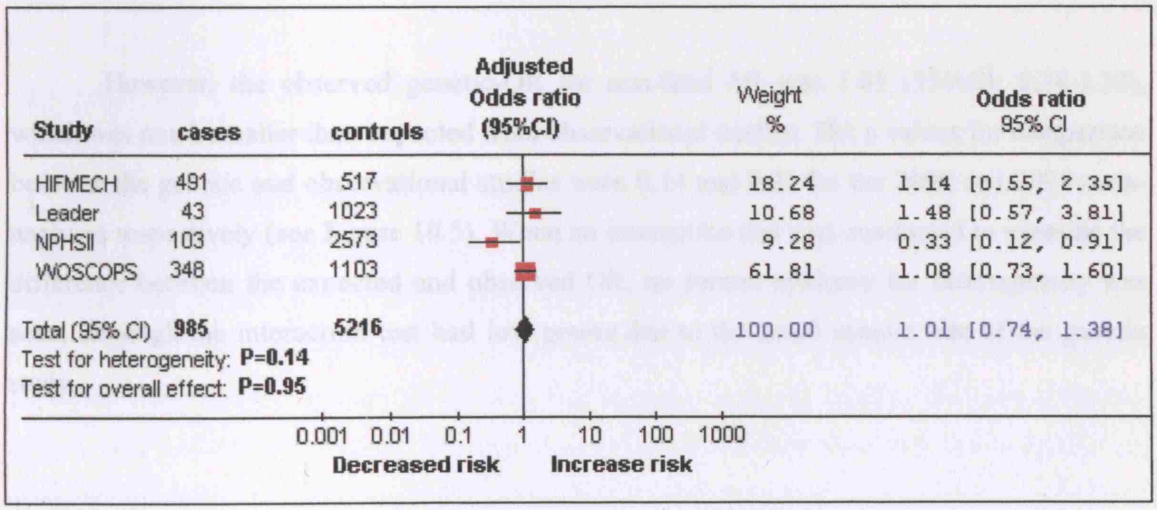
Study	Number of CHD cases	Observed relative risk* (95%CI)	Expected risk in genetic studies† (95%CI)
Meta-analysis 2000	1953	2.00 (1.60-2.50)	1.33 (1.09-1.67)
Meta-analysis 2004	7068	1.58 (1.48-1.68)	1.21 (1.06-1.37)

*Top vs. bottom tertile of CRP (corresponding to around 2.4 vs. 1.0mg/L difference in CRP).
†Assuming the CRP-CHD association is causal, with a log-linear relationship between CRP and CHD, and equivalence between relative risk and odds ratio. The confidence intervals take into account the uncertainty in both observational and genetic associations through the generation of simulated distributions (undertaken by Dr. Juan Pablo Casas and Dr. Leonelo Bautista).

10.4.5 Relationship between the +1444C/T genotype and IHD (non-fatal MI)

Data from four studies involving 985 cases (subjects with non-fatal MI) and 5216 control subjects were pooled in a meta-analysis to obtain a summary adjusted-OR. None of the four individual odds ratios (adjusted or unadjusted) indicated a significant increase in risk of non-fatal MI for TT-homozygous individuals compared with C-allele carriers. After combining the studies under a fixed effect model, TT-homozygous individuals had no significant increase in the risk of non-fatal MI (summary adjusted-OR=1.01, 95%CI: 0.74-1.38, p=0.95) compared with C-allele carriers (see Figure 10.3). No significant inter-study heterogeneity was observed (p=0.14). When the risk of non-fatal MI was restricted to the prospective studies, there was still no significant association (OR=0.98, 95%CI: 0.70-1.38, p=0.93). Similar results were also observed under a co-dominant model of inheritance, which gave a risk of non-fatal MI of 0.93 (95%CI: 0.67-1.28, p=0.65) for TT-homozygous individuals compared to CC-homozygous individuals.

Figure 10.3. Adjusted odds ratio* for non-fatal MI among subjects with the +1444TT-genotype in comparison with carriers of the +1444C-allele.



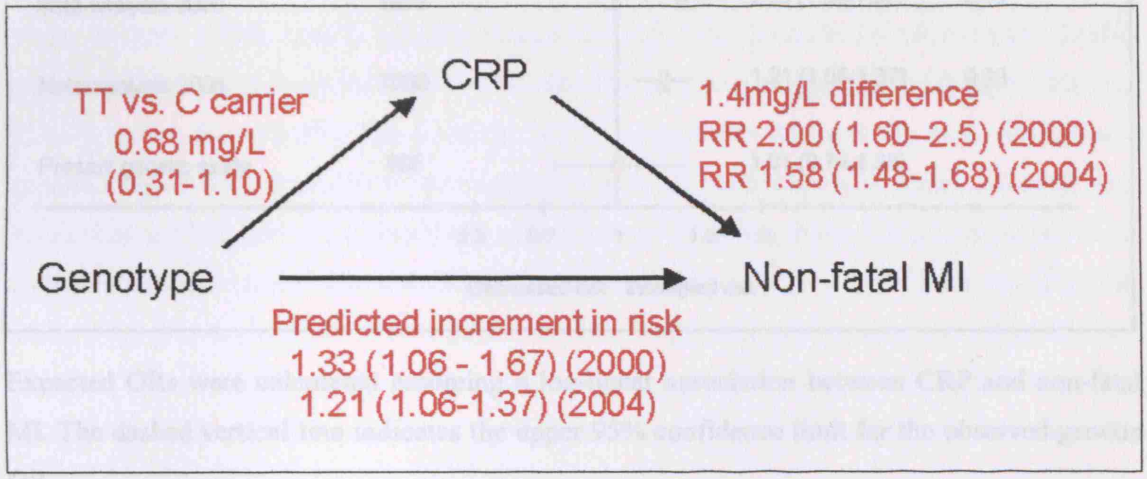
*Adjusted for age, hypertension, total-cholesterol, body mass index, diabetes mellitus, alcohol intake and smoking status.

10.4.6 Consistency between observed and expected odds ratios

Consistency between the expected and the observed OR was evaluated assuming a linear relationship between log-CRP and the risk of coronary events and using a recessive genetic model. The expected OR for non-fatal MI was derived from the meta-analysis of the effect of CRP genotype on CRP concentration and the previously published summary risks from two meta-analyses of prospective studies, which gave a predicted effect of 1.33 (95%CI: 1.09-

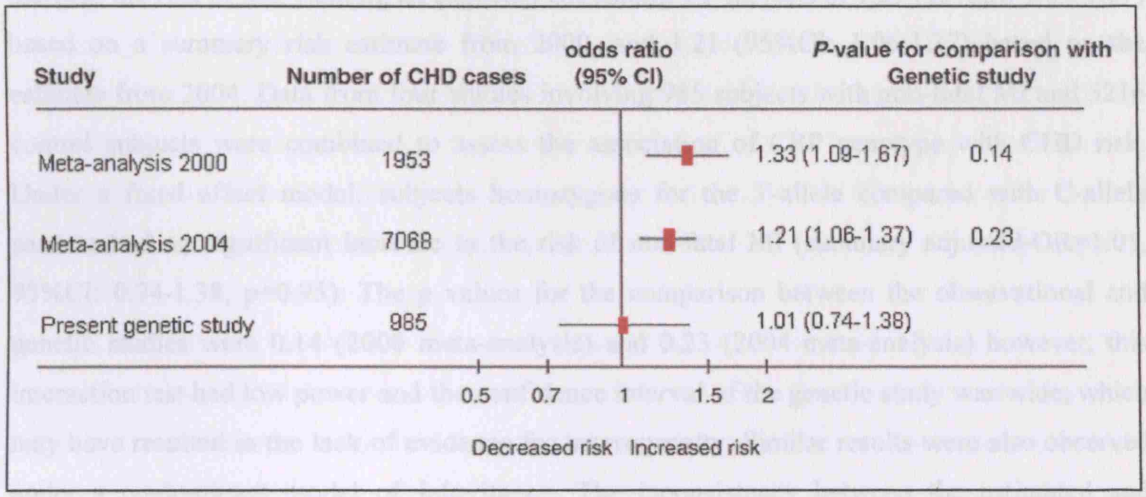
1.67) using the summary risk from the meta-analysis in 2000, and 1.21 (95%CI: 1.06-1.37) from the meta-analysis in 2004.

Figure 10.4. Triangulation of expected odds ratio for non-fatal MI for individuals homozygous for the CRP +1444C/T polymorphism.



However, the observed genetic-OR for non-fatal MI was 1.01 (95%CI: 0.74-1.38), which was much smaller than expected from observational studies. The p values for comparison between the genetic and observational studies were 0.14 and 0.23 for the 2000 and 2004 meta-analyses respectively (see Figure 10.5). When an interaction test was conducted to examine the difference between the expected and observed OR, no formal evidence for heterogeneity was seen, although the interaction test had low power due to the small sample size of the genetic study.

Figure 10.5. Comparison of expected and observed odds ratios for non-fatal MI for individuals homozygous for the CRP +1444C/T polymorphism.



Expected ORs were calculated assuming a log-linear association between CRP and non-fatal MI. The dashed vertical line indicates the upper 95% confidence limit for the observed genetic OR.

10.5 Discussion

Several prospective studies have demonstrated that higher than average CRP concentrations among apparently healthy individuals are associated with higher risk for future cardiovascular events (Ridker & Haughie 1998; Ridker *et al.* 2002; Danesh & Pepys 2000; Danesh *et al.* 1998). This has led to the suggestion that CRP may be causal in coronary disease. However, observational studies are not able to determine causality and the association may be confounded or subject to reverse causality. Since randomised controlled trials are currently not possible, an alternative approach is to use Mendelian randomisation.

This study found that the +1444C/T polymorphism (rs1130864) was reliably associated with CRP concentration. Using data from 4659 individuals (6 data sets), men homozygous for the +1444T allele had a CRP concentration 0.68 mg/L (95%CI: 0.31-1.10), higher than C-allele carriers. This difference was independent of cardiovascular risk factors or other indices of the inflammatory response such as fibrinogen across all studies and was similar to the lack of association seen between the +1444C/T SNP and covariables in the NPHSII cohort in Chapter 7. The randomised allocation of alleles appeared to balance the distribution of covariates in all data sets and thus minimise confounding. Therefore, CRP SNPs such as the +1444C/T variant appear to be a less biased estimate on coronary risk than CRP itself and can thus be used as a proxy to investigate the association with non-fatal MI risk.

If CRP were causal, a 0.68 mg/L increment in CRP concentration would be expected to increase the risk of MI, yielding an expected OR among TT subjects of 1.33 (95%CI: 1.09-1.67) based on a summary risk estimate from 2000, and 1.21 (95%CI: 1.06-1.37) based on the estimate from 2004. Data from four studies involving 985 subjects with non-fatal MI and 5216 control subjects were combined to assess the association of CRP genotype with CHD risk. Under a fixed effect model, subjects homozygous for the T-allele compared with C-allele carriers had no significant increase in the risk of non-fatal MI (summary adjusted-OR=1.01, 95%CI: 0.74-1.38, p=0.95). The p values for the comparison between the observational and genetic studies were 0.14 (2000 meta-analysis) and 0.23 (2004 meta-analysis) however, this interaction test had low power and the confidence interval of the genetic study was wide, which may have resulted in the lack of evidence for heterogeneity. Similar results were also observed under a co-dominant model of inheritance. The inconsistency between the estimated and observed risk of non-fatal MI, among subjects homozygous for this variant, suggests that the association between CRP concentration and later coronary events in previous prospective studies may be subject to residual confounding, reverse causality bias or both.

However, the confidence intervals for the estimated effect and the actual genetic effect seen were wide, with the upper limit of the genetic effect being almost 1.4 and the lower limit of the estimated effect being around 1.1. Therefore, despite having studied nearly 1000 cases and about 4500 controls, there was still some overlap of the confidence limits between the observed and expected OR, and thus not possible to completely exclude a small but potentially important causal influence of CRP on CHD on a population scale. In addition, since data were only available from around 1000 cases, the study only had 30% power to detect an odds ratio of 1.21 among TT subjects for non-fatal MI at α=0.05. The problem of accessing a data set large enough can be overcome by pooling data from many large studies to yield a combined data set of around 20000 cases and controls that would be required to detect an OR as low as 1.1 using a genotype with a frequency of around 0.2 with 90% power at α=0.01 (see Figure 10.1), and a network of collaborators is currently being established for such a task. A complementary goal is to better capture the phenotypic variation through the use of haplotypes, or SNPs with a larger effect on the CRP gene (including identification of functional SNPs), in order to better estimate the effect of CRP on coronary risk as this might reduce the number of cases required for such analyses.

Several assumptions underlie the Mendelian randomisation approach utilised here. These include the random distribution of alleles at conception, leading to the balancing of potential confounders between genotypic classes. This work showed that risk factors and other biomarkers were evenly distributed among the +1444 C-allele carriers and TT-homozygous

individuals, thereby minimising the potential for confounding. Mendelian randomisation also assumes that the difference in CRP concentration by genotype is constant throughout the life course. Although this assumption was not addressed in this work, it could be tested by examining the gene-CRP association in individuals at various times in the life course. Another assumption is that the difference in CRP by genotype is similar through strata of other covariables. This assumption was explored in work in Chapter 7, in which there was no evidence for a modification of the effect of genotype or haplotype on CRP concentration. CRP genotype had a similar proportional effect on CRP at all levels of age, BMI, blood pressure etc.

Several polymorphisms have now been studied in the CRP gene, where both associations with CRP concentration and associations with disease have been examined. Published studies and in house data sets have robustly shown association between several SNPs in the CRP gene and CRP concentration (see Chapters 6 and 7). However, fewer studies have assessed the relationship between CRP SNPs and coronary disease. Published studies reported that a +1059G/C polymorphism in the CRP gene is associated with an approximately 0.3-0.4 mg/L difference in CRP concentration, but not with arterial thrombosis or venous thromboembolism (Zee & Ridker 2002). Similar null associations were also seen between this polymorphism and hypertension (Davey Smith *et al.* 2005b). In addition, several SNPs including the –286C/T/A, +194A/T, +1059G/C, +1444C/T and +2302G/A all showed null association with risk of incident MI or ischaemic stroke in a prospective nested case-control study (Miller *et al.* 2005).

Recently, there has been more focus on tagging SNPs and association studies utilising haplotypes generated from these tagging SNPs. One advantage of using haplotypes over single SNPs is that most, if not all the variation along the gene will be captured, which is useful if the causal variant(s) is unknown. However, there will be more haplotypic groups generated than genotypic groups (which are only three for a biallelic variant), resulting in a loss of power that could generate false positive and false negative data. This would mean that a larger number of cases and controls would be required to try and detect association. In the case of Mendelian randomisation, it is not necessary to have the causal variant, as long as the polymorphism in question is associated reliably with plasma biomarker concentration, and therefore may be in LD with the causal variant. Recent haplotype studies combining the +1444C/T variant with other tagging SNPs have also seen null associations with metabolic syndrome and coronary disease (Timpson *et al.* 2005; Kardys *et al.* 2006), suggesting the work in this chapter is in line with published data.

Although randomised controlled trials are the most preferred method of determining causality and testing reversibility, they are not always possible. In the case of CRP, no selective

drug is currently available for such trials, although one is in the early stages of development (Pepys *et al.* 2006). Mendelian randomisation analysis may help prioritise biomarkers associated with coronary disease to be studied as drug targets, perhaps helping with the efficacy of new drug development.

10.6 Conclusions

This study verified the association between the +1444C/T genotype in CRP and plasma concentrations, where individuals homozygous for the T-allele had a weighted mean difference in concentrations of 0.68mg/L (95%CI 0.31-1.10) compared to C-allele carriers. By evaluating the distribution of other cardiovascular risk factors such as age, blood pressure, cholesterol, smoking and fibrinogen, it was possible to confirm that the association between genotype and CRP levels was not subject to confounding by these risk factors. When the risk of non-fatal MI in TT-homozygous individuals was compared with the risk in C-allele carriers, no significant difference was seen. Therefore, Caucasian men with a genotype that would have exposed them to a long-term elevation in CRP concentration were not at increased risk of non-fatal MI (OR: 1.01; 95% CI 0.74-1.38). However, to exclude any potential causal role of CRP in the pathogenesis of coronary disease, large genetic association studies that are sufficiently powered are required. Taking into account that the estimates of the risk of CRP concentrations on CHD are still uncertain and attenuating over time, it is possible to assume that unbiased and non-confounded estimates of the effect of CRP on coronary events are smaller than previously estimated.

Discussion and conclusions

11. Discussion and overall conclusions

Cardiovascular disease is one of the main causes of premature death in the developed world. Although a number of risk factors have been established from population studies, experimental studies and randomised controlled trials, there remains interest that additional risk factors might exist. It is thought that inflammatory processes contribute to the initiation and development of atherosclerotic lesions, as well as the ultimate development of acute ischaemic syndromes. There has therefore been more focussed interest on the role of specific markers of inflammation and their potential causal role in atherosclerosis and as screening tests to help identify those at higher risk of disease. Notable amongst these markers are fibrinogen, interleukin-6 (IL-6), serum amyloid A (SAA) and C-reactive protein (CRP), which have been consistently associated in observational studies with later cardiovascular events.

CRP is a major acute phase protein that displays a marked rise of its plasma concentration in response to infection or inflammation. Twin studies have suggested a significant degree of heritability (0.22-0.54) of circulating CRP concentration in health, providing support for an important genetic contribution to regulation. Prospective studies have shown that higher than average baseline CRP concentrations among apparently healthy individuals are associated with an increased risk for future cardiovascular events (Ridker *et al.* 2002; Danesh *et al.* 2000a; Danesh *et al.* 2004). This has led to the proposals that CRP measurement may be useful in the prediction of cardiovascular events and may play a causal role in atherosclerosis.

The first part of this thesis addressed the issue of the utility of CRP in the detection of later coronary heart disease events in initially healthy middle-aged men from the NPHSII study, comparing its performance with other risk factors and biomarkers, and assessing whether CRP adds useful predictive information over and above traditional risk factors and the Framingham risk score. The association between CRP and coronary events was similar in magnitude to that observed in numerous previous observational studies (Shlipak *et al.* 2005; Tzoulaki *et al.* 2005; St Pierre *et al.* 2005; Cushman *et al.* 2005; Koenig *et al.* 2004; Danesh *et al.* 2004; Curb *et al.* 2003; Sakkinen *et al.* 2002; Folsom *et al.* 2002; Lowe 2001; Harris *et al.* 1999; Ridker *et al.* 1997). However, when the performance of CRP in the detection of coronary events was assessed by measuring detection rates, false positive rates and constructing receiver operating characteristic (ROC) curves, it appeared to be a poor discriminator of subjects who do or do not develop coronary disease. When CRP was added to the Framingham risk score to see if it might provide additional predictive value, there was minimal improvement in the area under the ROC curve, showing CRP added little to the predictive performance. Re-evaluation of published data indicated substantial consistency of NPHSII data with those from other data sets.

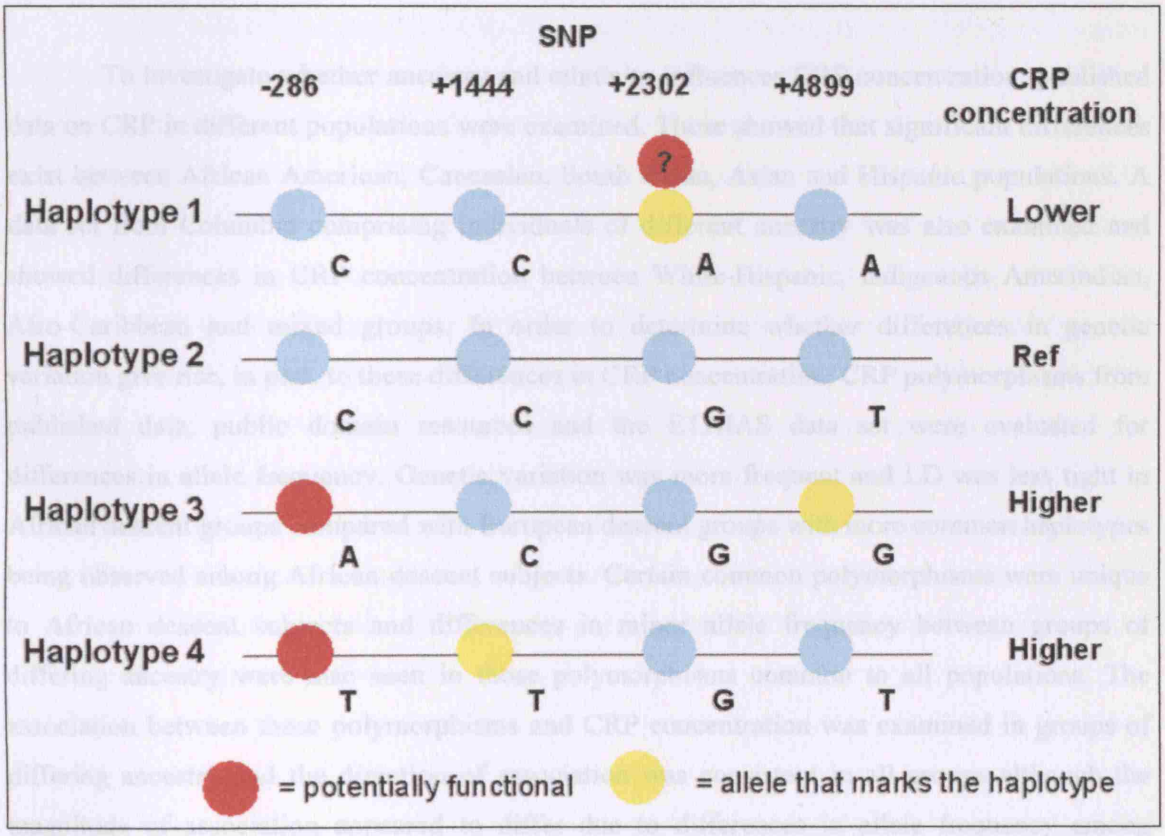
Since the predictive utility of a CRP measure may be limited, the association of CRP with cardiovascular disease was evaluated for its potential causal relevance. If CRP plays a pathogenic role in cardiovascular disease risk, then it could be an important therapeutic target and its reduction might help CVD prevention. However, it is only possible to make limited inference on causation from observational data because of the problems of confounding and reverse causality. Studying CRP indirectly via polymorphisms that influence its concentration may provide a better insight on causation since genotype is allocated at random at conception, thus reducing confounding and overcoming reverse causality. Therefore, a consensus genetic map containing 31 variants within and flanking the CRP gene was constructed by collating and comparing information from public domain databases of sequence variation. Linkage disequilibrium (LD) within the gene was also examined, allowing tagging SNPs to be generated. For a European population, it was determined that just three tagging SNPs would be required to capture more than 90% of the variation in the CRP gene, and for an African population, six tagging SNPs would be required. These panels of common SNPs could therefore be used to conduct genotype-intermediate phenotype and genotype-disease association studies.

In order to determine whether CRP polymorphisms were associated with CRP concentration, meta-analyses of published data from 12 different studies were conducted to provide an overview of association. Six common polymorphisms were assessed, from which 5 showed association with CRP concentration. The minor alleles of the –286C/T/A (rs3091244), +194A/T (rs1417938), and +1444C/T (rs1130864) polymorphisms were all associated with higher CRP concentration. The minor alleles of the +1059G/C (rs1800947) and +2302G/A (rs1205) polymorphisms were associated with lower CRP concentration. The +1444C/T and +2302G/A polymorphisms are two of the CRP tagging SNPs and define different haplotypes. Therefore, since they are both strongly associated with CRP concentration and are not in LD, it suggests the presence of more than one causal variant in the CRP gene.

Associations were retested in additional data sets to confirm the association and allow more precise estimates of the genotype-CRP association to be made. Four polymorphisms were assessed, comprising the three tagging SNPs required for European populations (+1444C/T, +2302G/A, +4899T/G), and one potentially functional variant (-286C/T/A). In addition, common CRP haplotypes were inferred by genotyping tagging SNPs, to allow tests for association of haplotype with CRP concentration. The minor alleles of the –286C/T/A, +1444C/T and +4899T/G polymorphisms were individually all associated with higher CRP concentration in the NPHSII and Ely data sets. The +1444T-allele marks the –286T-allele and the +4899G-allele marks the –286A-allele. Therefore, it is possible that the –286 site contains two functional variants in European populations (see Figure 11.1). The minor allele of the

+2302G/A polymorphism was associated with lower CRP concentration, and this SNP, which lies on a distinct haplotype, is not in LD with any other common known SNPs in the immediate vicinity. Therefore, this SNP may be a causal variant itself or may be involved in a regulatory site some distance away from the CRP gene. These results were consistent with the associations with CRP concentration seen in published data. Haplotypes generated from the three tagging SNPs were also robustly associated with CRP concentration. The effect of genotype and haplotype on CRP concentration did not appear to be confounded by other variables also associated with CRP concentration, even when the data were stratified by quantiles of covariates. They were also consistent with haplotype analyses from published work.

Figure 11.1. Haplotypes defined by three tagging SNPs in European populations and location of potentially functional SNPs.



Four common haplotypes are generated from the +1444C/T, +2302G/A and +4899T/G SNPs. The –286C/T/A SNP, which may be functional, is in LD with the +1444T (T-allele) and +4899G (A-allele) SNPs and lies on these two haplotypes (Haplotypes 3 and 4). The +2302G/A SNP, which defines Haplotype 1, is not in LD with known common variants in the CRP gene and may therefore be functional itself or may be in LD with an unknown functional variant up or downstream of the CRP gene.

Since the associations between these polymorphisms and haplotypes and CRP concentration thus far had been seen in healthy individuals, the associations were also assessed during acute inflammation in individuals with periodontal disease. As the population of interest had a mixed ethnicity, six tagging SNPs were evaluated, both individually and in haplotypes. All CRP SNPs were associated with differences in CRP concentration during the acute phase response produced by periodontal treatment, with the exception of the –717A/G polymorphism. In addition, the ten haplotypes generated from the tagging SNPs were also associated with peak CRP concentration and the results were concordant with data from analyses in the absence of acute inflammation, demonstrating that CRP is subject to genetic modulation over a wide range of concentrations seen in both health and disease. This effect was independent of conventional cardiovascular risk factors and inflammatory factors (such as IL-6) that are known to affect CRP concentrations, suggesting that high CRP concentrations seen in acute inflammatory conditions and infections may not be solely due to increased levels of inflammation.

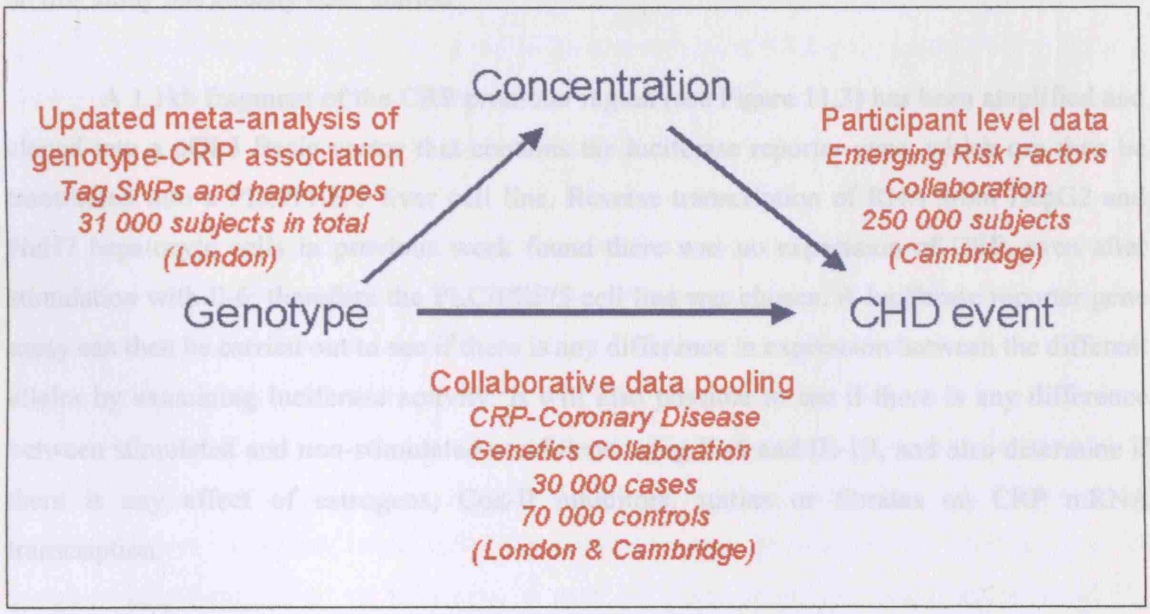
To investigate whether ancestry and ethnicity influences CRP concentration, published data on CRP in different populations were examined. These showed that significant differences exist between African American, Caucasian, South Asian, Asian and Hispanic populations. A data set from Columbia comprising individuals of different ancestry was also examined and showed differences in CRP concentration between White-Hispanic, indigenous Amerindian, Afro-Caribbean and mixed groups. In order to determine whether differences in genetic variation give rise, in part, to these differences in CRP concentration, CRP polymorphisms from published data, public domain resources and the ETNIAS data set were evaluated for differences in allele frequency. Genetic variation was more frequent and LD was less tight in African descent groups compared with European descent groups with more common haplotypes being observed among African descent subjects. Certain common polymorphisms were unique to African descent subjects and differences in minor allele frequency between groups of differing ancestry were also seen in those polymorphisms common to all populations. The association between these polymorphisms and CRP concentration was examined in groups of differing ancestry and the direction of association was consistent in all groups although the magnitude of association appeared to differ due to differences in allele frequency among ancestral groups. This work may have importance for the clinical utility of CRP as a predictive tool since CRP concentration appears to be influenced by ancestral background, and CRP polymorphisms that also affect CRP concentration exhibit altered frequency by ancestry.

The last part of this thesis aimed to determine whether CRP has causal relevance in coronary disease utilising the natural randomisation of alleles at conception in a Mendelian randomisation approach that should minimise biases such as confounding and reverse causality. The +1444C/T was chosen for this analysis and its effect on CRP concentration and also on

non-fatal MI was evaluated. Using data from 4659 men, those homozygous for the +1444T-allele had a CRP concentration 0.68 mg/L higher than C-allele carriers, which was independent of cardiovascular risk factors and biomarkers across all studies and was similar to the lack of association seen between the +1444C/T SNP and covariables in the NPHSII and Ely data sets in Chapter 7. In order to estimate the expected odds ratio for non-fatal MI, data from prior non-genetic observational studies examining the CRP-coronary event association were used. Data from 985 subjects with non-fatal MI and 5216 control subjects were then combined to determine the association between genotype and disease. This showed that men homozygous for the T-allele compared with C-allele carriers had no significant increase in the risk of non-fatal MI. However, the confidence intervals for the estimated effect and the actual genetic effect seen were wide, therefore, it was not possible to completely exclude a small to moderate causal influence of CRP on CHD, which could still be of importance for coronary prevention.

Power calculations conducted by Professor Liam Smeeth (London School of Hygiene and Tropical Medicine) as part of the ongoing CRP-CHD Genetics Collaboration, designed to determine the number of individuals required to detect a genetic effect size that could still correspond to a causal effect of CRP have suggested that sample sizes of around 20 000 CHD cases and a similar number of controls are required. Larger, collaborative Mendelian randomisation analyses are now underway to evaluate this definitively (see Figure 11.2). The Emerging Risk Factors Collaboration (Professor John Danesh, University of Cambridge) will provide individual participant data on the association between CRP concentration and CHD. The meta-analyses of published data and in house data sets have already been undertaken to assess the association between genotypes and haplotypes and CRP concentration. The final part of the Mendelian randomisation analyses will involve the CRP-CHD Genetics Collaboration (coordinated by Dr. Aroon Hingorani and Professor John Danesh) with data on around 30 000 cases and 70 000 controls, which will be adequately powered to definitively determine whether CRP is causal in coronary disease.

Figure 11.2. Collaborations underway to evaluate the role of CRP in CHD.



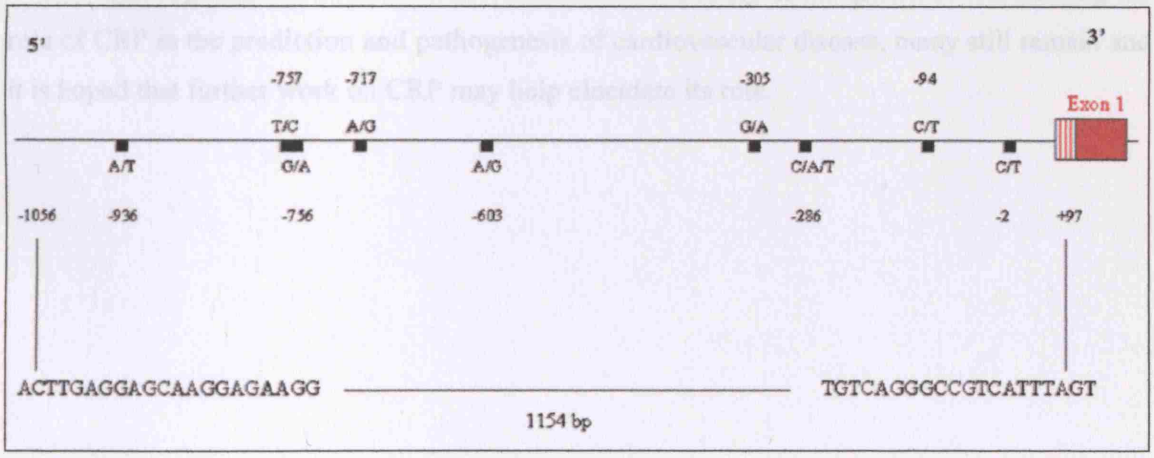
Since CRP displays a marked rise of its serum concentration in response to infection or inflammation, it is widely used in clinical practice as a marker of progress and response to treatment in infective or inflammatory processes. Studies in this thesis showed that polymorphisms in the CRP gene are associated with CRP concentration in the presence of acute inflammation. This indicates that CRP genotype or haplotype could affect the prognostic or diagnostic abilities of CRP as a marker for infection or inflammation. Studies are currently underway in acute coronary syndrome cohorts to evaluate this.

As yet, the variant that regulates CRP concentration cannot be ascribed with certainty, and there is a strong possibility that more than one functional polymorphism exists, distributed over more than one haplotype. Work could be conducted to assess whether the promoter variants associated with differences in CRP concentration could be functional, by investigating their influence on CRP transcription *in vitro*, both basally and during cytokine stimulation with IL-6 and IL-1β. It would also be possible to investigate if there is any effect of estrogens, Cox-II inhibitors, statins or fibrates on mRNA transcription, as all of these have been associated with differences in CRP concentration in population studies, and in many cases, the effects could be independent of an alteration of inflammatory burden on CRP. The –305G/A and –286C/T/A polymorphisms both lie within E-box transcription factor binding sites and may be putatively functional. Alternatively, they may be in linkage disequilibrium with other variants downstream or upstream of the CRP gene that are causal, so association seen between these two promoter variants and CRP concentration may be confounded by LD. Individuals with the homozygous forms of the common and rare allele(s) for each polymorphism can be compared to investigate

if there are any differences in rates of mRNA transcription, in order to resolve this issue. Work on this study has already been started.

A 1.1kb fragment of the CRP promoter region (see Figure 11.3) has been amplified and cloned into a pGL3 Basic vector that contains the luciferase reporter gene, which can then be transfected into a PLC/PRF/5 liver cell line. Reverse transcription of RNA from HepG2 and HuH7 hepatocyte cells in previous work found there was no expression of CRP, even after stimulation with Il-6; therefore the PLC/PRF/5 cell line was chosen. A luciferase reporter gene assay can then be carried out to see if there is any difference in expression between the different alleles by examining luciferase activity. It will also possible to see if there is any difference between stimulated and non-stimulated conditions using IL-6 and IL-1β, and also determine if there is any effect of estrogens, Cox-II inhibitors, statins or fibrates on CRP mRNA transcription.

Figure 11.3. Region of CRP promoter being amplified for promoter-reporter constructs, with the sequence of the primer pair and its location shown below. All promoter polymorphisms are shown, along with their location (bases). The 5' UTR and Exon 1 are also shown in red.



In addition, the +2302G/A polymorphism lies on a distinct haplotype and work in this thesis has shown that this variant is strongly associated with CRP concentration. This SNP does not appear to be in LD with other known common variants in the CRP gene, suggesting the +2302G/A SNP may be functional. Since this site lies in the 3' UTR, it may be situated in or near instability elements such as AU-rich elements (AREs) and may therefore be involved in mRNA stability (Xu *et al.* 1997). Less is known about regulatory elements in the 3' UTR region and not as many bioinformatic tools are available for assessment compared to promoter regions. Experimental, biochemical assays could be utilised to investigate *in vitro* whether the

+2302G/A SNP is involved in degradation of CRP mRNA (Fritz *et al.* 2000). *In vivo* analyses may also be possible using a HaploChIP approach. This technique allows the relative allele-specific transcriptional activity of a gene to be assayed using RNA polymerase II (Pol II) as a surrogate marker, and uses mass spectrometry to quantify differences in activity (Knight *et al.* 2003).

Several avenues of further exploration have been exposed by this work. The chapter on ancestry showed that several polymorphisms associated with CRP concentration and the response to inflammation are preserved in several populations of differing genetic background. One potential explanation for this preservation is that they protect the individual from infection, since CRP is able to recognise pathogens by binding phosphocholine and mediates their elimination by recruiting the complement system and phagocytic cells. It is possible that infection drives evolution more than late onset diseases. By examining whether rates of common infections differ by genotype, it may be possible to understand more about the ancestral origins of these polymorphisms and their role in infections. Genotype work can also be carried out in archived samples from Africa to assess this.

Although this thesis has identified and answered many of the questions surrounding the role of CRP in the prediction and pathogenesis of cardiovascular disease, many still remain and it is hoped that further work on CRP may help elucidate its role.

References

12. References

Agerholm-Larsen B, Nordestgaard BG, Steffensen R, Sorensen TIA, Jensen G & Tybjµrg-Hansen A 1997 ACE Gene Polymorphism: Ischemic Heart Disease and Longevity in 10 150 Individuals: A Case-Referent and Retrospective Cohort Study Based on the Copenhagen City Heart Study. *Circulation* 95 2358-2367.

Agerholm-Larsen B, Nordestgaard BG & Tybjarg-Hansen A 2000 ACE Gene Polymorphism in Cardiovascular Disease: Meta-Analyses of Small and Large Studies in Whites. *Arteriosclerosis, Thrombosis, and Vascular Biology* 20 484-492.

Agmon Y, Khandheria BK, Meissner I, Schwartz GL, Petterson TM, O'Fallon WM, Gentile F, Whisnant JP, Wiebers DO & Seward JB 2000 Independent association of high blood pressure and aortic atherosclerosis: A population-based study. *Circulation* 102 2087-2093.

Agrawal A, Cha-Molstad H, Samols D & Kushner I 2001 Transactivation of C-reactive protein by IL-6 requires synergistic interaction of CCAAT/enhancer binding protein β (C/EBPβ) and Rel p50. *Journal of Immunology* 166 2378-2384.

Agrawal A, Samols D & Kushner I 2003 Transcription factor c-Rel enhances C-reactive protein expression by facilitating the binding of C/EBPβ to the promoter. *Molecular Immunology* 40 373-380.

Airila-Mansson S, Soder B, Jin LJ, Soder PO & Klinge B 2004 Self-reporting of periodontal diseases and clinical assessment outcome in a Swedish urban population of smokers and non-smokers. *Acta Odontol.Scand.* 62 111-115.

Aitkenhead AR 1993 Postoperative ischaemia: cardiac morbidity after non-cardiac surgery. *Lancet* 341 731-732.

Alber DG, Powell KL, Vallance P, Goodwin DA & Grahame-Clarke C 2000 Herpesvirus Infection Accelerates Atherosclerosis in the Apolipoprotein E-Deficient Mouse. *Circulation* 102 779-785.

Albert MA, Glynn RJ, Buring J & Ridker PM 2004 C-Reactive protein levels among women of various ethnic groups living in the United States (from the Women's Health Study). *The American Journal of Cardiology* 93 1238-1242.

Albert MA & Ridker PM 2006 C-Reactive Protein as a Risk Predictor: Do Race/Ethnicity and Gender Make a Difference? *Circulation* 114 e67-e74.

Alley DE, Seeman TE, Ki Kim J, Karlamangla A, Hu P & Crimmins EM 2006 Socioeconomic status and C-reactive protein levels in the US population: NHANES IV. *Brain, Behavior, and Immunity* 20 498-504.

Allocco D, Song Q, Gibbons G, Ramoni M & Kohane I 2007 Geography and genography: prediction of continental origin using randomly selected single nucleotide polymorphisms. *BMC Genomics* 8 68.

Almirall J, Bolibar I, Toran P, Pera G, Boquet X, Balanzo X & Sauca G 2004 Contribution of C-Reactive Protein to the Diagnosis and Assessment of Severity of Community-Acquired Pneumonia. *Chest* 125 1335-1342.

Altman DG & Bland JM 1994a Statistics Notes: Diagnostic tests 1: sensitivity and specificity. *BMJ* 308 1552.

Altman DG & Bland JM 1994b Statistics Notes: Diagnostic tests 2: predictive values. *BMJ* 309 102.

Altman DG & Bland JM 2003 Statistics Notes: Interaction revisited: the difference between two estimates. *BMJ* 326 219.

Ambrose JA & Barua RS 2004 The pathophysiology of cigarette smoking and cardiovascular disease: An update. *Journal of the American College of Cardiology* 43 1731-1737.

Anand SS, Razak F, Davis AD, Jacobs R, Vuksan V, Teo K & Yusuf S 2006 Social disadvantage and cardiovascular disease: development of an index and analysis of age, sex, and ethnicity effects. *International Journal of Epidemiology* 35 1239-1245.

Anand SS, Razak F, Yi Q, Davis B, Jacobs R, Vuksan V, Lonn E, Teo K, McQueen M & Yusuf S 2004 C-Reactive Protein as a Screening Test for Cardiovascular Risk in a Multiethnic Population. *Arteriosclerosis, Thrombosis, and Vascular Biology* 24 1509-1515.

Anderson JL & Muhlestein JB 2000 The ACADEMIC study in perspective (Azithromycin in coronary artery disease: elimination of myocardial infection with Chlamydia). *J Infect.Dis.* 181 Suppl 3 S569-S571.

Anderson KM, Odell PM, Wilson PW & Kannel WB 1991 Cardiovascular disease risk profiles. *Am Heart J* 121 293-298.

Araujo F, Antelmi I, Pereira AC, Latorre MdR, Grupi CJ, Krieger JE & Mansur AJ 2006 Lower heart rate variability is associated with higher serum high-sensitivity C-reactive protein concentration in healthy individuals aged 46 years or more. *International Journal of Cardiology* 107 333-337.

Arcone R, Gualandi G & Ciliberto G 1988 Identification of sequences responsible for acute-phase induction of human C-reactive protein. *Nucleic Acids Res* 16 3195-3207.

Arita M, Ueno Y, Mohara O, Nakamura H, Tomobuchi Y, Yamamoto K, Nishio I & Masuyama Y 1987 The cause of death, risk factors and long-term prognosis of acute myocardial infarction. *Jpn.Circ.J* 51 332-337.

Assmann G, Cullen P & Schulte H 2002 Simple Scoring Scheme for Calculating the Risk of Acute Coronary Events Based on the 10-Year Follow-Up of the Prospective Cardiovascular Munster (PROCAM) Study. *Circulation* 105 310-315.

Austin MA, King MC, Bawol RD, Hulley SB & Friedman GD 1987 Risk factors for coronary heart disease in adult female twins. Genetic heritability and shared environmental influences. *Am J Epidemiol.* 125 308-318.

Austin MA, Hutter CM, Zimmern RL & Humphries SE 2004 Genetic Causes of Monogenic Heterozygous Familial Hypercholesterolemia: A HuGE Prevalence Review. *American Journal of Epidemiology* 160 407-420.

Balk EM, Lau J, Goudas LC, Jordan HS, Kupelnick B, Kim LU & Karas RH 2003 Effects of Statins on Nonlipid Serum Markers Associated with Cardiovascular Disease: A Systematic Review. *Annals of Internal Medicine* 139 670-682.

Barath P, Fishbein MC, Cao J, Berenson J, Helfant RH & Forrester JS 1990 Detection and localization of tumor necrosis factor in human atheroma. *Am.J.Cardiol.* 65 297-302.

Barinas-Mitchell E, Cushman M, Meilahn EN, Tracy RP & Kuller LH 2001 Serum Levels of C-reactive Protein Are Associated with Obesity, Weight Gain, and Hormone Replacement Therapy in Healthy Postmenopausal Women. *American Journal of Epidemiology* 153 1094-1101.

Barker DJP 2004 The Developmental Origins of Adult Disease. *Journal of the American College of Nutrition* 23 588S-595.

Barrett-Connor E & Khaw K 1984 Family history of heart attack as an independent predictor of death due to cardiovascular disease. *Circulation* 69 1065-1069.

Batty GD & Leon DA 2002 Socio-economic position and coronary heart disease risk factors in children and young people: Evidence from UK epidemiological studies. *The European Journal of Public Health* 12 263-272.

Baum L, Tomlinson B & Thomas GN 2003 APOA5-1131T>C polymorphism is associated with triglyceride levels in Chinese men. *Clinical Genetics* 63 377-379.

Berk B & Alexander R 1995 Biology of the vascular wall in hypertension. In *Pathophysiology of Renal Disease*, Ed Brenner BM. Philadelphia: WB Saunders.

Berk BC, Weintraub WS & Alexander RW 1990 Elevation of C-reactive protein in "active" coronary artery disease. *The American Journal of Cardiology* 65 168-172.

Bhagat K, Moss R, Collier J & Vallance P 1996 Endothelial "stunning" following a brief exposure to endotoxin: a mechanism to link infection and infarction? *Cardiovascular Research* 32 822-829.

Bhakdi S 2003 An hypothesis for the immunopathogenesis of atherosclerosis. *Clin.Nephrol.* 60 Suppl 1 S49-S52.

Bharadwaj D, Stein MP, Volzer M, Mold C & Du Clos TW 1999 The major receptor for C-reactive protein on leukocytes is fcγ receptor II. *The Journal of Experimental Medicine* 190 585-590.

Biasucci LM, Colizzi C, Rizzello V, Vitrella G, Crea F & Liuzzo G 1999a Role of inflammation in the pathogenesis of unstable coronary artery diseases. *Scand.J.Clin.Lab Invest Suppl* 230 12-22.

Biasucci LM, Liuzzo G, Buffon A & Maseri A 1999b The variable role of inflammation in acute coronary syndromes and in restenosis. *Semin.Interv.Cardiol.* 4 105-110.

Biasucci LM, Liuzzo G, Fantuzzi G, Caligiuri G, Rebuzzi AG, Ginnetti F, Dinarello CA & Maseri A 1999c Increasing Levels of Interleukin (IL)-1Ra and IL-6 During the First 2 Days of Hospitalization in Unstable Angina Are Associated With Increased Risk of In-Hospital Coronary Events. *Circulation* 99 2079-2084.

Biasucci LM, Liuzzo G, Grillo RL, Caligiuri G, Rebuzzi AG, Buffon A, Summaria F, Ginnetti F, Fadda G & Maseri A 1999d Elevated Levels of C-Reactive Protein at Discharge in Patients With Unstable Angina Predict Recurrent Instability. *Circulation* 99 855-860.

Bisoendial RJ, Kastelein JJP, Levels JHM, Zwaginga JJ, van den Bogaard B, Reitsma PH, Meijers JCM, Hartman D, Levi M & Stroes ESG 2005 Activation of Inflammation and Coagulation After Infusion of C-Reactive Protein in Humans. *Circulation Research* 96 714-716.

Blake GJ, Rifai N, Buring JE & Ridker PM 2003 Blood Pressure, C-Reactive Protein, and Risk of Future Cardiovascular Events. *Circulation* 108 2993-2999.

Blankenberg S, Rupprecht HJ, Bickel C, Peetz D, Hafner G, Tiret L & Meyer J 2001 Circulating cell adhesion molecules and death in patients with coronary artery disease. *Circulation* 104 1336-1342.

Blankenberg S, Barbaux S & Tiret L 2003 Adhesion molecules and atherosclerosis. *Atherosclerosis* 170 191-203.

Boekholdt SM, Hack CE, Sandhu MS, Luben R, Bingham SA, Wareham NJ, Peters RJG, Jukema JW, Day NE, Kastelein JJP & Khaw KT 2005 C-reactive protein levels and coronary artery disease incidence and mortality in apparently healthy men and women: The EPIC-Norfolk prospective population study 1993-2003. *Atherosclerosis* 187 415-422.

Boekholdt SM & Thompson JF 2003 Natural genetic variation as a tool in understanding the role of CETP in lipid levels and disease. *Journal of Lipid Research* 44 1080-1093.

Boer JM, Feskens EJ, Verschuren WM, Seidell JC & Kromhout D 1999 The joint impact of family history of myocardial infarction and other risk factors on 12-year coronary heart disease mortality. *Epidemiology* 10 767-770.

Boren J, Lee I, Zhu W, Arnold K, Taylor S & Innerarity TL 1998 Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its binding activity by the carboxyl terminus in familial defective apo-B100. *J Clin Invest* 101 1084-1093.

Borzak S & Ridker PM 1995 Discordance between Meta-analyses and Large-Scale Randomized, Controlled Trials: Examples from the Management of Acute Myocardial Infarction. *Annals of Internal Medicine* 123 873-877.

Braunwald E 1997 Cardiovascular Medicine at the Turn of the Millennium: Triumphs, Concerns, and Opportunities. *The New England Journal of Medicine* 337 1360-1369.

Bravata DM, Wells CK, Gulanski B, Kernan WN, Brass LM, Long J & Concato J 2005 Racial Disparities in Stroke Risk Factors: The Impact of Socioeconomic Status. *Stroke* 36 1507-1511.

Brindle P, Beswick A, Fahey T & Ebrahim S 2006 Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review. *Heart* 92 1752-1759.

Brindle P, Emberson J, Lampe F, Walker M, Whincup P, Fahey T & Ebrahim S 2003 Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study. *BMJ* 327 1267.

Brull DJ, Serrano N, Zito F, Jones L, Montgomery HE, Rumley A, Sharma P, Lowe GD, World MJ, Humphries SE & Hingorani AD 2003 Human CRP Gene Polymorphism Influences CRP Levels: Implications for the Prediction and Pathogenesis of Coronary Heart Disease. *Arteriosclerosis, Thrombosis, and Vascular Biology* 23 2063-2069.

Brunner EJ, Marmot MG, Nanchahal K, Shipley MJ, Stansfeld SA, Juneja M & Alberti KGMM 1997 Social inequality in coronary risk: central obesity and the metabolic syndrome. Evidence from the Whitehall II study. *Diabetologia* 40 1341-1349.

Burnett MS, Durrani S, Stabile E, Saji M, Lee CW, Kinnaird TD, Hoffman EP & Epstein SE 2004 Murine cytomegalovirus infection increases aortic expression of proatherosclerotic genes. *Circulation* 109 893-897.

Calver A, Collier J, Moncada S & Vallance P 1992 Effect of local intra-arterial NG-monomethyl-L-arginine in patients with hypertension: the nitric oxide dilator mechanism appears abnormal. *J Hypertens.* 10 1025-1031.

Cameron J, Holla OL, Ranheim T, Kulseth MA, Berge KE & Leren TP 2006 Effect of mutations in the PCSK9 gene on the cell surface LDL receptors. *Human Molecular Genetics* 15 1551-1558.

Cao H & Hegele RA 2000 Human C-reactive protein (CRP) 1059G/C polymorphism. *J.Hum.Genet.* 45 100-101.

Cao JJ, Thach C, Manolio TA, Psaty BM, Kuller LH, Chaves PHM, Polak JF, Sutton-Tyrrell K, Herrington DM, Price TR & Cushman M 2003 C-Reactive Protein, Carotid Intima-Media Thickness, and Incidence of Ischemic Stroke in the Elderly: The Cardiovascular Health Study. *Circulation* 108 166-170.

Cao Z, Umek RM & McKnight SL 1991 Regulated expression of three C/EBP isoforms during adipose conversion of 3T3-L1 cells. *Genes Dev.* 5 1538-1552.

Cappelleri JC, Ioannidis JP, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC & Lau J 1996 Large trials vs meta-analysis of smaller trials: how do their results compare? *JAMA: The Journal of the American Medical Association* 276 1332-1338.

Cappuccio FP, Cook DG, Atkinson RW & Strazzullo P 1997 Prevalence, detection, and management of cardiovascular risk factors in different ethnic groups in south London. *Heart* 78 555-563.

Cardia G, Grisorio D, Impedovo G, Lillo A & Regina G 1990 Plasma lipids as a risk factor in peripheral vascular disease. *Angiology* 41 19-22.

Carlson CS, Eberle MA, Kruglyak L & Nickerson DA 2004 Mapping complex disease loci in whole-genome association studies. *Nature* 429 446-452.

Carlson CS, Aldred SF, Lee PK, Tracy RP, Schwartz SM, Rieder M, Liu K, Williams OD, Iribarren C, Lewis EC, Fornage M, Boerwinkle E, Gross M, Jaquish C, Nickerson DA, Myers RM, Siscovick DS & Reiner AP 2005 Polymorphisms within the C-Reactive Protein (CRP) Promoter Region Are Associated with Plasma CRP Levels. *The American Journal of Human Genetics* 77 64-77.

Casas JP, Bautista LE, Smeeth L, Sharma P & Hingorani AD 2005 Homocysteine and stroke: evidence on a causal link from mendelian randomisation. *The Lancet* 365 224-232.

Cave AC, Brewer AC, Narayanapanicker A, Ray R, Grieve DJ, Walker S & Shah AM 2006 NADPH Oxidases in Cardiovascular Health and Disease. *Antioxidants & Redox Signaling* 8 691-728.

Celermajer DS, Adams MR, Clarkson P, Robinson J, McCredie R, Donald A & Deanfield JE 1996 Passive smoking and impaired endothelium-dependent arterial dilatation in healthy young adults. *The New England Journal of Medicine* 334 150-154.

Celermajer DS, Sorensen KE, Georgakopoulos D, Bull C, Thomas O, Robinson J & Deanfield JE 1993 Cigarette smoking is associated with dose-related and potentially reversible impairment of endothelium-dependent dilation in healthy young adults. *Circulation* 88 2149-2155.

Cha-Molstad H, Agrawal A, Zhang D, Samols D & Kushner I 2000 The Rel family member P50 mediates cytokine-induced C-reactive protein expression by a novel mechanism. *Journal of Immunology* 165 4592-4597.

Chambers JC, Eda S, Bassett P, Karim Y, Thompson SG, Gallimore JR, Pepys MB & Kooner JS 2001 C-Reactive Protein, Insulin Resistance, Central Obesity, and Coronary Heart Disease Risk in Indian Asians From the United Kingdom Compared With European Whites. *Circulation* 104 145-150.

Chandalia M, Cabo-Chan AV, Jr., Devaraj S, Jialal I, Grundy SM & Abate N 2003 Elevated Plasma High-Sensitivity C-Reactive Protein Concentrations in Asian Indians Living in the United States. *Journal of Clinical Endocrinology Metabolism* 88 3773-3776.

Chatha K, Anderson NR & Gama R 2002a Ethnic variation in C-reactive protein: UK resident Indo-Asians compared with Caucasians. *Journal of Cardiovascular Risk* 9 139-141.

Chen CJ, Cohen BH, Diamond EL, Lin TM & Chen JS 1984 Genetic variance and heritability of cardiovascular risk factors in Chinese adolescent twins. *Acta Genet Med Gemellol.(Roma.)* 33 363-373.

Ciliberto G, Arcone R, Wagner EF & Ruther U 1987 Inducible and tissue-specific expression of human C-reactive protein in transgenic mice. *The EMBO Journal* 6 4017-4022.

Colhoun HM, Rubens MB, Underwood SR & Fuller JH 2000 The effect of type 1 diabetes mellitus on the gender difference in coronary artery calcification. *J.Am.Coll.Cardiol.* 36 2160-2167.

Colhoun HM, McKeigue PM & Davey Smith G 2003 Problems of reporting genetic associations with complex outcomes. *The Lancet* 361 865-872.

Collins FS, Guyer MS & Chakravarti A 1997 Variations on a Theme: Cataloging Human DNA Sequence Variation. *Science* 278 1580-1581.

Collins R, Peto R, MacMahon S, Hebert P, Fiebach NH, Eberlein KA, Godwin J, Qizilbash N, Taylor JO & Hennekens CH 1990 Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* 335 827-838.

Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA & Pritchard JK 2006 A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet* 38 1251-1260.

Cooper RS, Kaufman JS & Ward R 2003 Race and Genomics. *The New England Journal of Medicine* 348 1166-1170.

Corre S & Galibert MD 2005 Upstream stimulating factors: highly versatile stress-responsive transcription factors. *Pigment Cell Research* 18 337-348.

Craig SR, Leaver HA, Yap PL, Pugh GC & Walker WS 2001 Acute phase responses following minimal access and conventional thoracic surgery. *European Journal of Cardio-Thoracic Surgery* 20 455-463.

Crawford D, Yi Q, Smith J, Shephard C, Wong M, Witrak L, Livingston R, Rieder M & Nickerson D 2006 Allelic spectrum of the natural variation in CRP. *Human Genetics* 119 496-504.

Crawford DC & Nickerson DA 2005 Definition and clinical importance of haplotypes. *Annual Review of Medicine* 56 303-320.

Curb JD, Abbott RD, Rodriguez BL, Sakkinen P, Popper JS, Yano K & Tracy RP 2003 C-reactive protein and the future risk of thromboembolic stroke in healthy men. *Circulation* 107 2016-2020.

Cushman M, Meilahn E, Psaty B, Kuller L, Dobs A & Tracy R 1999a Hormone Replacement Therapy, Inflammation, and Hemostasis in Elderly Women. *Arteriosclerosis, Thrombosis, and Vascular Biology* 19 893-899.

Cushman M, Arnold AM, Psaty BM, Manolio TA, Kuller LH, Burke GL, Polak JF & Tracy RP 2005 C-Reactive Protein and the 10-Year Incidence of Coronary Heart Disease in Older Men and Women: The Cardiovascular Health Study. *Circulation* 112 25-31.

Cushman M, Legault C, Barrett-Connor E, Stefanick ML, Kessler C, Judd HL, Sakkinen PA & Tracy RP 1999b Effect of Postmenopausal Hormones on Inflammation-Sensitive Proteins : The Postmenopausal Estrogen/Progestin Interventions (PEPI) Study. *Circulation* 100 717-722.

D'Aiuto F, Casas JP, Shah T, Humphries SE, Hingorani AD & Tonetti MS 2005 C-reactive protein (+1444C>T) polymorphism influences CRP response following a moderate inflammatory stimulus. *Atherosclerosis* 179 413-417.

Dahlen G, Ericson C, de Faire U, Iselius L & Lundman T 1983 Genetic and environmental determinants of cholesterol and HDL-cholesterol concentrations in blood. *Int.J Epidemiol.* 12 32-35.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ & Lander ES 2001 High-resolution haplotype structure in the human genome. *Nat Genet* 29 229-232.

Danesh J & Pepys MB 2000 C-reactive protein in healthy and in sick populations. *European Heart Journal* 21 1564-1565.

Danesh J, Whincup P, Walker M, Lennon L, Thomson A, Appleby P, Gallimore JR & Pepys MB 2000a Low grade inflammation and coronary heart disease: prospective study and updated meta-analyses. *BMJ* 321 199-204.

Danesh J 2005 Antibiotics in the prevention of heart attacks. *The Lancet* 365 365-367.

Danesh J, Collins R, Appleby P & Peto R 1998 Association of Fibrinogen, C-reactive Protein, Albumin, or Leukocyte Count With Coronary Heart Disease: Meta-analyses of Prospective Studies. *JAMA: The Journal of the American Medical Association* 279 1477-1482.

Danesh J, Collins R & Peto R 1997 Chronic infections and coronary heart disease: is there a link? *The Lancet* 350 430-436.

Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, Lowe GDO, Pepys MB & Gudnason V 2004 C-Reactive Protein and Other Circulating Markers of Inflammation in the Prediction of Coronary Heart Disease. *The New England Journal of Medicine* 350 1387-1397.

Danesh J, Whincup P, Walker M, Lennon L, Thomson A, Appleby P, Wong Yk, Bernardes-Silva M, Ward M & West R 2000b Chlamydia pneumoniae IgG titres and coronary heart disease: prospective study and meta-analysis Commentary: Adjustment for potential confounders may have been taken too far. *BMJ* 321 208-213.

Davey Smith G & Ebrahim S 2003 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* 32 1-22.

Davey Smith G & Ebrahim S 2004 Mendelian randomization: prospects, potentials, and limitations. *International Journal of Epidemiology* 33 30-42.

Davey Smith G & Egger M 1998 Incommunicable Knowledge? Interpreting and Applying the Results of Clinical Trials and Meta-Analyses. *Journal of Clinical Epidemiology* 51 289-295.

Davey Smith G, Harbord R, Milton J, Ebrahim S & Sterne JAC 2005a Does Elevated Plasma Fibrinogen Increase the Risk of Coronary Heart Disease?: Evidence from a Meta-Analysis of Genetic Association Studies. *Arteriosclerosis, Thrombosis, and Vascular Biology* 25 2228-2233.

Davey Smith G, Lawlor DA, Harbord R, Timpson N, Rumley A, Lowe GDO, Day INM & Ebrahim S 2005b Association of C-Reactive Protein With Blood Pressure and Hypertension: Life Course

Confounding and Mendelian Randomization Tests of Causality. *Arteriosclerosis, Thrombosis & Vascular Biology* 25 1051-1056.

Davies JR, Rudd JF, Fryer TD & Weissberg PL 2005 Targeting the vulnerable plaque: The evolving role of nuclear imaging. *Journal of Nuclear Cardiology* 12 234-246.

De Backer G, Ambrosioni E, Broch-Johnsen K, Brotons C, Cifkova R, Dallongeville J, Ebrahim S, Faergeman O, Graham I, Mancia G, Cats VM, Orth-Gomer K, Perk J, Pyorala K, Rodicio J, Sans S, Sansoy V, Sechtem U, Silber S, Thomsen T & Wood D 2003 European guidelines on cardiovascular disease prevention in clinical practice Third Joint Task Force of European and other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of eight societies and by invited experts). *European Journal of Cardiovascular Prevention & Rehabilitation* 10 S1-S78.

De Backer J, Mak R, De Bacquer D, Van Renterghem L, Verbraekel E, Kornitzer M & De Backer G 2002 Parameters of inflammation and infection in a community based case-control study of coronary heart disease. *Atherosclerosis* 160 457-463.

de Bakker PIW, Burtt NP, Graham RR, Guiducci C, Yelensky R, Drake JA, Bersaglieri T, Penney KL, Butler J, Young S, Onofrio RC, Lyon HN, Stram DO, Haiman CA, Freedman ML, Zhu X, Cooper R, Groop L, Kolonel LN, Henderson BE, Daly MJ, Hirschhorn JN & Altshuler D 2006 Transferability of tag SNPs in genetic association studies in multiple populations. *Nat Genet* 38 1298-1303.

de Maat MPM, Bladbjerg EM, Hjelmborg JvB, Bathum L, Jespersen J & Christensen K 2004 Genetic Influence on Inflammation Variables in the Elderly. *Arteriosclerosis, Thrombosis & Vascular Biology* 24 2168-2173.

Deeb S, Failor A, Brown BG, Brunzell JD, Albers JJ & Motulsky AG 1986 Molecular genetics of apolipoproteins and coronary heart disease. *Cold Spring Harb.Symp.Quant.Biol* 51 Pt 1 403-409.

Deeks JJ & Altman DG 2004 Diagnostic tests 4: likelihood ratios. *BMJ* 329 168-169.

DerSimonian R & Laird N 1986 Meta-analysis in clinical trials. *Control Clin.Trials* 7 177-188.

Deshpande RG, Khan M & Genco CA 1998 Invasion strategies of the oral pathogen porphyromonas gingivalis: implications for cardiovascular disease. *Invasion Metastasis* 18 57-69.

DeStefano F, Anda RF, Kahn HS, Williamson DF & Russell CM 1993 Dental disease and risk of coronary heart disease and mortality. *BMJ* 306 688-691.

Devaraj S, Xu DY & Jialal I 2003 C-reactive protein increases plasminogen activator inhibitor-1 expression and activity in human aortic endothelial cells: implications for the metabolic syndrome and atherothrombosis. *Circulation* 107 398-404.

Dhamrait SS, Stephens JW, Cooper JA, Acharya J, Mani AR, Moore K, Miller GJ, Humphries SE, Hurel SJ & Montgomery HE 2004 Cardiovascular risk in healthy men and markers of oxidative stress in diabetic men are associated with common variation in the gene for uncoupling protein 2. *European Heart Journal* 25 468-475.

Diez Roux AV, Auchincloss AH, Astor B, Barr RG, Cushman M, Dvonch T, Jacobs DR, Jr., Kaufman J, Lin X & Samson P 2006 Recent Exposure to Particulate Matter and C-reactive Protein Concentration in the Multi-Ethnic Study of Atherosclerosis. *American Journal of Epidemiology* 164 437-448.

Douglas JA, Erdos MR, Watanabe RM, Braun A, Johnston CL, Oeth P, Mohlke KL, Valle TT, Ehnholm C, Buchanan TA, Bergman RN, Collins FS, Boehnke M & Tuomilehto J 2001 The Peroxisome

Poliferator-Activated Receptor-ã2 Pro12Ala Variant: Association With Type 2 Diabetes and Trait Differences. *Diabetes* 50 886-890.

Duncan BB, Schmidt MI, Chambless LE, Folsom AR, Carpenter M & Heiss G 2000 Fibrinogen, Other Putative Markers of Inflammation, and Weight Gain in Middle-aged Adults--The ARIC Study. *Obesity* 8 279-286.

Ebrahim S, Sung J, Song YM, Ferrer RL, Lawlor DA & Smith GD 2006 Serum cholesterol, haemorrhagic stroke, ischaemic stroke, and myocardial infarction: Korean national health system prospective cohort study. *BMJ* 333 22.

Eklund C, Lehtimaki T & Hurme M 2005 Epistatic effect of C-reactive protein (CRP) single nucleotide polymorphism (SNP) +1059 and interleukin-1B SNP +3954 on CRP concentration in healthy male blood donors. *International Journal of Immunogenetics* 32 229-232.

Eriksson JG, Forsen T, Tuomilehto J, Osmond C & Barker DJP 2001 Early growth and coronary heart disease in later life: longitudinal study. *BMJ* 322 949-953.

Eriksson JG 2005 The fetal origins hypothesis--10 years on. *BMJ* 330 1096-1097.

Fabricant CG, Fabricant J, Litrenta MM & Minick CR 1978 Virus-induced atherosclerosis. *The Journal of Experimental Medicine* 148 335-340.

Fan J & Watanabe T 2003 Inflammatory reactions in the pathogenesis of atherosclerosis. *J Atheroscler.Thromb* 10 63-71.

Farese RV, Linton MF & Young SG 1992 Apolipoprotein B gene mutations affecting cholesterol levels. *J Intern Med* 231 643-652.

Farrall M, Green FR, Peden JF, Olsson PG, Clarke R, Hellenius ML, Rust S, Lagercrantz J, Franzosi MG, Schulte H, Carey A, Olsson G, Assmann G, Tognoni G, Collins R, Hamsten A & Watkins H 2006 Genome-Wide Mapping of Susceptibility to Coronary Artery Disease Identifies a Novel Replicated Locus on Chromosome 17. *PLoS Genetics* 2 e72.

Festa A, D'Agostino R, Jr., Williams K, Karter AJ, Mayer-Davis EJ, Tracy RP & Haffner SM 2001 The relation of body fat mass and distribution to markers of chronic inflammation. *Int.J.Obes.Relat Metab Disord.* 25 1407-1415.

Fibrinogen studies Collaboration 2005 Plasma Fibrinogen Level and the Risk of Major Cardiovascular Diseases and Nonvascular Mortality: An Individual Participant Meta-analysis. *JAMA: The Journal of the American Medical Association* 294 1799-1809.

Fiuza C & Suffredini AF 2001 Human models of innate immunity: local and systemic inflammatory responses. *J Endotoxin Res* 7 385-388.

Folsom AR, Aleksic N, Catellier D, Juneja HS & Wu KK 2002 C-reactive protein and incident coronary heart disease in the Atherosclerosis Risk In Communities (ARIC) study. *American Heart Journal* 144 233-238.

Folsom AR, Chambless LE, Ballantyne CM, Coresh J, Heiss G, Wu KK, Boerwinkle E, Mosley TH, Jr., Sorlie P, Diao G & Sharrett AR 2006 An Assessment of Incremental Coronary Risk Prediction Using C-Reactive Protein and Other Novel Risk Markers: The Atherosclerosis Risk in Communities Study. *Archives of Internal Medicine* 166 1368-1373.

Folsom AR, Pankow JS, Tracy RP, Arnett DK, Peacock JM, Hong Y, Djousse L & Eckfeldt JH 2001 Association of C-reactive protein with markers of prevalent atherosclerotic disease. *The American Journal of Cardiology* 88 112-117.

Fong IW, Chiu B, Viira E, Fong MW, Jang D & Mahony J 1997 Rabbit model for Chlamydia pneumoniae infection. *J Clin Microbiol.* 35 48-52.

Fong IW, Chiu B, Viira E, Jang D, Fong MW, Peeling R & Mahony JB 1999 Can an antibiotic (macrolide) prevent Chlamydia pneumoniae-induced atherosclerosis in a rabbit model? *Clin Diagn.Lab Immunol.* 6 891-894.

Ford ES 1999 Body mass index, diabetes, and C-reactive protein among U.S. adults. *Diabetes Care* 22 1971-1977.

Ford ES, Giles WH, Mokdad AH & Myers GL 2004 Distribution and Correlates of C-Reactive Protein Concentrations among Adult US Women. *Clinical Chemistry* 50 574-581.

Ford ES, Giles WH, Myers GL & Mannino DM 2003a Population Distribution of High-Sensitivity C-reactive Protein among US Men: Findings from National Health and Nutrition Examination Survey 1999-2000. *Clinical Chemistry* 49 686-690.

Ford ES, Giles WH, Myers GL, Rifai N, Ridker PM & Mannino DM 2003b C-reactive Protein Concentration Distribution among US Children and Young Adults: Findings from the National Health and Nutrition Examination Survey, 1999-2000. *Clinical Chemistry* 49 1353-1357.

Forouhi N, Sattar N & McKeigue P 2001 Relation of C-reactive protein to body fat distribution and features of the metabolic syndrome in Europeans and South Asians. *International Journal of Obesity* 25 1327-1331.

Forouhi NG & Sattar N 2006 CVD risk factors and ethnicity--A homogeneous relationship? *Atherosclerosis Supplements* 7 11-19.

Foussas SG, Zairis MN, Lyras AG, Patsourakos NG, Tsirimpis VG, Katsaros K, Beldekos DJ, Handanis SM, Mytas DZ & Karidis KS 2005 Early Prognostic Usefulness of C-Reactive Protein Added to the Thrombolysis In Myocardial Infarction Risk Score in Acute Coronary Syndromes. *The American Journal of Cardiology* 96 533-537.

Frankel S, Elwood P, Smith GD, Frankel S, Sweetnam P & Yarnell J 1996 Birthweight, body-mass index in middle age, and incident coronary heart disease. *The Lancet* 348 1478-1480.

Fransen EJ, Maessen JG, Hermens W, Glatz JFC & Buurman WA 2000 Peri-operative myocardial tissue injury and the release of inflammatory mediators in coronary artery bypass graft patients. *Cardiovascular Research* 45 853-859.

Friedlander Y, Austin MA, Newman B, Edwards K, Mayer-Davis EJ & King MC 1997 Heritability of Longitudinal Changes in Coronary-Heart-Disease Risk Factors in Women Twins. *The American Journal of Human Genetics* 60 1502-1512.

Friedlander Y, Kark JD & Stein Y 1985 Family history of myocardial infarction as an independent risk factor for coronary heart disease. *Br.Heart J* 53 382-387.

Friedlander Y, Kark JD & Stein Y 1986 Biological and environmental sources of variation in plasma lipids and lipoproteins: the Jerusalem Lipid Research Clinic. *Hum.Hered.* 36 143-153.

Fritz DT, Ford LP & Wilusz J 2000 An in Vitro Assay to Study Regulated mRNA Stability. *Science's STKE* 2000 11.

Frohlich M, Sund M, Thorand B, Hutchinson WL, Pepys MB & Koenig W 2002 Lack of Seasonal Variation in C-Reactive Protein. *Clinical Chemistry* 48 575-577.

Fulop AK, Pocsik E, Brozik M, Karabelyos C, Kiss A, Novak I, Szalai C, Dobozy O & Falus A 2001 Hepatic regeneration induces transient acute phase reaction: systemic elevation of acute phase reactants and soluble cytokine receptors. *Cell Biology International* 25 585-592.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ & Altshuler D 2002 The Structure of Haplotype Blocks in the Human Genome. *Science* 296 2225-2229.

Galvani M, Ferrini D, Ghezzi F & Ottani F 2001 Cardiac markers and risk stratification: an integrated approach. *Clinica Chimica Acta* 311 9-17.

Ganapathi MK, Rzewnicki D, Samols D, Jiang SL & Kushner I 1991 Effect of combinations of cytokines and hormones on synthesis of serum amyloid A and C-reactive protein in Hep 3B cells. *J.Immunol.* 147 1261-1265.

Gattone M, Iacoviello L, Colombo M, Castelnuovo AD, Soffiantino F, Gramoni A, Picco D, Benedetta M & Giannuzzi P 2001 Chlamydia pneumoniae and cytomegalovirus seropositivity, inflammatory markers, and the risk of myocardial infarction at a young age*1. *American Heart Journal* 142 633-640.

George J, Harats D, Gilburd B, Levy Y, Langevitz P & Shoenfeld Y 1999 Atherosclerosis-related markers in systemic lupus erythematosus patients: the role of humoral immunity in enhanced atherogenesis. *Lupus* 8 220-226.

Gibson FC, III, Hong C, Chou HH, Yumoto H, Chen J, Lien E, Wong J & Genco CA 2004 Innate immune recognition of invasive bacteria accelerates atherosclerosis in apolipoprotein E-deficient mice. *Circulation* 109 2801-2806.

Goff DC, Nichaman MZ, Chan W, Ramsey DJ, Labarthe DR & Ortiz C 1997 Greater Incidence of Hospitalized Myocardial Infarction Among Mexican Americans Than Non-Hispanic Whites : The Corpus Christi Heart Project, 1988-1992. *Circulation* 95 1433-1440.

Goldman ND, Liu T & Lei KJ 1987 Structural analysis of the locus containing the human C-reactive protein gene and its related pseudogene. *Journal of Biological Chemistry* 262 7001-7005.

Gordon PA, George J, Khamashta MA, Harats D, Hughes G & Shoenfeld Y 2001 Atherosclerosis and autoimmunity. *Lupus* 10 249-252.

Gordon T, Kannel WB, Castelli WP & Dawber TR 1981 Lipoproteins, cardiovascular disease, and death. The Framingham study. *Archives of Internal Medicine* 141 1128-1131.

Grayston JT, Kuo CC, Campbell LA, Wang SP & Jackson LA 1997 Chlamydia pneumoniae and cardiovascular disease. *Cardiologia (Rome, Italy)* 42 1145-1151.

Greenfield JR, Samaras K, Jenkins AB, Kelly PJ, Spector TD, Gallimore JR, Pepys MB & Campbell LV 2004 Obesity Is an Important Determinant of Baseline Serum C-Reactive Protein Concentration in Monozygotic Twins, Independent of Genetic Influences. *Circulation* 109 3022-3028.

Griendling KK & Alexander RW 1997 Oxidative stress and cardiovascular disease. *Circulation* 96 3264-3265.

Grimes DS, Hindle E & Dyer T 2000 Respiratory infection and coronary heart disease: progression of a paradigm. *QJM* 93 375-383.

Gu L, Okada Y, Clinton SK, Gerard C, Sukhova GK, Libby P & Rollins BJ 1998 Absence of monocyte chemoattractant protein-1 reduces atherosclerosis in low density lipoprotein receptor-deficient mice. *Mol.Cell* 2 275-281.

Gupta S, Leatham EW, Carrington D, Mendall MA, Kaski JC & Camm AJ 1997 Elevated Chlamydia pneumoniae antibodies, cardiovascular events, and azithromycin in male survivors of myocardial infarction. *Circulation* 96 404-407.

Hansson GK 2001 Regulation of immune mechanisms in atherosclerosis. *Ann.N Y.Acad.Sci.* 947 157-165.

Harland JO, Unwin N, Bhopal RS, White M, Watson B, Laker M & Alberti KG 1997 Low levels of cardiovascular risk factors and coronary heart disease in a UK Chinese population. *Journal of Epidemiology and Community Health* 51 636-642.

Harris TB, Ferrucci L, Tracy RP, Corti MC, Wacholder S, EttingerJr WH, Heimovitz H, Cohen HJ & Wallace R 1999 Associations of elevated Interleukin-6 and C-Reactive protein levels with mortality in the elderly. *The American Journal of Medicine* 106 506-512.

Hattori Y, Matsumura M & Kasai K 2003 Vascular smooth muscle cell activation by C-reactive protein. *Cardiovascular Research* 58 186-195.

Haverkate F, Thompson SG, Pyke SDM, Gallimore JR, Pepys MB & Concerted Action on Thrombosis and Disabilities Angina Pectoris Study Group E 1997 Production of C-reactive protein and risk of coronary events in stable and unstable angina. *The Lancet* 349 462-466.

Heald AH, Anderson SG, Ivison F, Laing I, Gibson JM & Cruickshank K 2003 C-reactive protein and the insulin-like growth factor (IGF)-system in relation to risk of cardiovascular disease in different ethnic groups. *Atherosclerosis* 170 79-86.

Heart Protection Study Collaborative G 2002 MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20[punctuation space]536 high-risk individuals: a randomised placebo-controlled trial. *The Lancet* 360 23-33.

Heeschen C, Hamm CW, Bruemmer J & Simoons ML 2000 Predictive value of C-reactive protein and troponin T in patients with unstable angina: a comparative analysis. *Journal of the American College of Cardiology* 35 1535-1542.

Heinrich PC, Castell JV & Andus T 1990 Interleukin-6 and the acute phase response. *Biochem.J.* 265 621-636.

Herrera VL, Shen L, Lopez LV, Didishvili T, Zhang YX & Ruiz-Opazo N 2003 Chlamydia pneumoniae accelerates coronary artery disease progression in transgenic hyperlipidemia-genetic hypertension rat model. *Mol.Med* 9 135-142.

Higgins JP, Thompson SG, Deeks JJ & Altman DG 2003 Measuring inconsistency in meta-analyses. *BMJ* 327 557-560.

Hirschfield GM & Pepys MB 2003 C-reactive protein and cardiovascular disease: new insights from an old molecule. *QJM.* 96 793-807.

Hobbs HH, Brown MS & Goldstein JL 1992 Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. *Hum.Mutat.* 1 445-466.

Hong SH, Song J & Kim JQ 2000 Genetic variations of the hepatic lipase gene in Korean patients with coronary artery disease. *Clinical Biochemistry* 33 291-296.

Horkko S, Binder CJ, Shaw PX, Chang MK, Silverman G, Palinski W & Witztum JL 2000 Immunological responses to oxidized LDL. *Free Radic.Biol Med* 28 1771-1779.

Hort W 1994 Arteriosclerosis: its morphology in the past and today. *Basic Res.Cardiol.* 89 Suppl 1 1-15.

Hort W 2002 History of cardiovascular pathology. *Z.Kardiol.* 91 Suppl 4 20-24.

Howard BV, Criqui MH, Curb JD, Rodabough R, Safford MM, Santoro N, Wilson AC & Wylie-Rosett J 2003 Risk factor clustering in the insulin resistance syndrome and its relationship to cardiovascular disease in postmenopausal white, black, hispanic, and Asian/Pacific Islander women. *Metabolism* 52 362-371.

Howard BV, Lee ET, Cowan LD, Devereux RB, Galloway JM, Go OT, Howard WJ, Rhoades ER, Robbins DC, Sievers ML & Welty TK 1999 Rising Tide of Cardiovascular Disease in American Indians : The Strong Heart Study. *Circulation* 99 2389-2395.

Hsich E, Zhou YF, Paigen B, Johnson TM, Burnett MS & Epstein SE 2001 Cytomegalovirus infection increases development of atherosclerosis in Apolipoprotein-E knockout mice. *Atherosclerosis* 156 23-28.

Humphries SE, Whittall RA, Hubbart CS, Maplebeck S, Cooper JA, Soutar A, Naoumova R, Thompson GR, Seed M, Durrington PN, Miller JP, Betteridge DJ & Neil HAW 2006 Genetic causes of Familial Hypercholesterolaemia in UK patients: relation to plasma lipid levels and coronary heart disease risk. *Journal of Medical Genetics* In press.

Hundt M, Zielinska-Skowronek M & Schmidt RE 2001 Lack of specific receptors for C-reactive protein on white blood cells. *Eur.J Immunol* 31 3475-3483.

Hunt KJ, Williams K, Resendez RG, Hazuda HP, Haffner SM & Stern MP 2002 All-Cause and Cardiovascular Mortality Among Diabetic Participants in the San Antonio Heart Study: Evidence against the "Hispanic Paradox". *Diabetes Care* 25 1557-1563.

Huxley R, Owen CG, Whincup PH, Cook DG, Colman S & Collins R 2004 Birth Weight and Subsequent Cholesterol Levels: Exploration of the "Fetal Origins" Hypothesis. *JAMA: The Journal of the American Medical Association* 292 2755-2764.

IJzerman RG, Stehouwer CDA, van Weissenbruch MM, de Geus EJ & Boomsma DI 2001 Evidence for Genetic Factors Explaining the Association Between Birth Weight and Low-Density Lipoprotein Cholesterol and Possible Intrauterine Factors Influencing the Association Between Birth Weight and High-Density Lipoprotein Cholesterol: Analysis in Twins. *Journal of Clinical Endocrinology Metabolism* 86 5479-5484.

Ikeda U, Takahashi M & Shimada K 1998 Monocyte-endothelial cell interaction in atherogenesis and thrombosis. *Clin Cardiol.* 21 11-14.

Imhof A, Frohlich M, Loewel H, Helbecque N, Woodward M, Amouyel P, Lowe GDO & Koenig W 2003 Distributions of C-reactive Protein Measured by High-Sensitivity Assays in Apparently Healthy Men and Women from Different Populations in Europe. *Clinical Chemistry* 49 669-672.

Inazu A, Jiang XC, Haraki T, Yagi K, Kamon N, Koizumi J, Mabuchi H, Takeda R, Takata K & Moriyama Y 1994 Genetic cholesteryl ester transfer protein deficiency caused by two prevalent mutations as a major determinant of increased levels of high density lipoprotein cholesterol. *J Clin Invest* 94 1872-1882.

Inazu A, Nishimura Y, Terada Y & Mabuchi H 2001 Effects of hepatic lipase gene promoter nucleotide variations on serum HDL cholesterol concentration in the general Japanese population. *Journal of Human Genetics* 46 172-177.

Ioannidis JP 2005 Molecular Bias. *European Journal of Epidemiology* 20 739-745.

Ioannidis JPA, Cappelleri JC & Lau J 1998 Issues in Comparisons Between Meta-analyses and Large Trials. *JAMA: The Journal of the American Medical Association* 279 1089-1093.

Ioannidis JPA, Gwinn M, Little J, Higgins JPT, Bernstein JL, Boffetta P, Bondy M, Bray MS, Brenchley PE, Buffler PA, Casas JP, Chokkalingam A, Danesh J, Smith GD, Dolan S, Duncan R, Gruis NA, Hartge P, Hashibe M, Hunter DJ, Jarvelin MR, Malmer B, Maraganore DM, Newton-Bishop JA, O'Brien TR, Petersen G, Riboli E, Salanti G, Seminara D, Smeeth L, Taioli E, Timpson N, Uitterlinden AG, Vineis P, Wareham N, Winn DM, Zimmern R, Khoury MJ & The Human Genome Epidemiology Network and the Network of Investigator Networks 2006 A road map for efficient and reliable human genome epidemiology. *Nat Genet* 38 3-5.

Ioannidis JP, Ntzani EE & Trikalinos TA 2004 'Racial' differences in genetic effects for complex diseases. *Nat Genet* 36 1312-1318.

Isshiki H, Akira S, Sugita T, Nishio Y, Hashimoto S, Pawlowski T, Suematsu S & Kishimoto T 1991 Reciprocal expression of NF-IL6 and C/EBP in hepatocytes: possible involvement of NF-IL6 in acute phase protein gene expression. *New Biol* 3 63-70.

Jackson R, Lawes CM, Bennett DA, Milne RJ & Rodgers A 2005 Treatment with drugs to lower blood pressure and blood cholesterol based on an individual's absolute cardiovascular risk. *The Lancet* 365 434-441.

Jacobsen BK & Thelle DS 1988 Risk factors for coronary heart disease and level of education: The Tromso Heart Study. *American Journal of Epidemiology* 127 923-932.

Jager A, van Hinsbergh VWM, Kostense PJ, Emeis JJ, Yudkin JS, Nijpels G, Dekker JM, Heine RJ, Bouter LM & Stehouwer CDA 1999 von Willebrand Factor, C-Reactive Protein, and 5-Year Mortality in Diabetic and Nondiabetic Subjects:The Hoorn Study. *Arteriosclerosis, Thrombosis, and Vascular Biology* 19 3071-3078.

James SK, Armstrong P, Barnathan E, Califf R, Lindahl B, Siegbahn A, Simoons ML, Topol EJ, Venge P & Wallentin L 2003 Troponin and C-reactive protein have different relations to subsequent mortality and myocardial infarction after acute coronary syndrome: A GUSTO-IV substudy. *Journal of the American College of Cardiology* 41 916-924.

Jansen H, Verhoeven AJM, Weeks L, Kastelein JJP, Halley DJJ, van den Ouweland A, Jukema JW, Seidell JC & Birkenhager JC 1997 Common C-to-T Substitution at Position -480 of the Hepatic Lipase Promoter Associated With a Lowered Lipase Activity in Coronary Artery Disease Patients. *Arteriosclerosis, Thrombosis, and Vascular Biology* 17 2837-2842.

Jeffreys AJ, Kauppi L & Neumann R 2001 Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* 29 217-222.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG & Todd JA 2001 Haplotype tagging for the identification of common disease genes. *Nat.Genet.* 29 233-237.

Jones DW, Chambless LE, Folsom AR, Heiss G, Hutchinson RG, Sharrett AR, Szklo M & Taylor HA, Jr. 2002 Risk Factors for Coronary Heart Disease in African Americans: The Atherosclerosis Risk in Communities Study, 1987-1997. *Archives of Internal Medicine* 162 2565-2571.

Jousilahti P, Salomaa V, Rasi V, Vahtera E & Palosuo T 2003 Association of markers of systemic inflammation, C reactive protein, serum amyloid A, and fibrinogen, with socioeconomic status. *Journal of Epidemiology and Community Health* 57 730-733.

Judson R, Brain C, Dain B, Windemuth A, Ruano G & Reed C 2004 New and confirmatory evidence of an association between APOE genotype and baseline C-reactive protein in dyslipidemic individuals. *Atherosclerosis* 177 345-351.

Juge-Aubry CE, Somm E, Giusti V, Pernin A, Chicheportiche R, Verdumo C, Rohner-Jeanrenaud F, Burger D, Dayer JM & Meier CA 2003 Adipose tissue is a major source of interleukin-1 receptor antagonist: upregulation in obesity and inflammation. *Diabetes* 52 1104-1110.

Juhan-Vague I, Morange PE, Aubert H, Henry M, Aillaud MF, Alessi MC, Samnegard A, Hawe E, Yudkin J, Margaglione M, Di Minno G, Hamsten A & Humphries SE 2002 Plasma thrombin-activatable fibrinolysis inhibitor antigen concentration and genotype in relation to myocardial infarction in the north and south of Europe. *Arterioscler.Thromb.Vasc.Biol.* 22 867-873.

Kadiri S, Walker O, Salako BL & Akinkugbe O 1999 Blood pressure, hypertension and correlates in urbanised workers in Ibadan, Nigeria: a revisit. *J Hum.Hypertens.* 13 23-27.

Kalin MF & Zumoff B 1990 Sex hormones and coronary disease: a review of the clinical studies. *Steroids* 55 330-352.

Kannel WB, D'Agostino RB & Belanger AJ 1987 Fibrinogen, cigarette smoking, and risk of cardiovascular disease: insights from the Framingham Study. *Am Heart J* 113 1006-1010.

Kannel WB & McGee DL 1979a Diabetes and cardiovascular disease. The Framingham study. *JAMA: The Journal of the American Medical Association* 241 2035-2038.

Kannel WB & McGee DL 1979b Diabetes and cardiovascular risk factors: the Framingham study. *Circulation* 59 8-13.

Kaplan MH & Volanakis JE 1974 Interaction of C reactive protein complexes with the complement system. I. Consumption of human complement associated with the reaction of C reactive protein with pneumococcal C polysaccharide and with the choline phosphatides, lecithin and sphingomyelin. *J.Immunol.* 112 2135-2147.

Kardys I, de Maat MPM, Uitterlinden AG, Hofman A & Witteman JCM 2006 C-reactive protein gene haplotypes and risk of coronary heart disease: the Rotterdam Study. *European Heart Journal* 27 1331-1337.

Kawanami D, Maemura K, Takeda N, Harada T, Nojiri T, Saito T, Manabe I, Imai Y & Nagai R 2006 C-reactive protein induces VCAM-1 gene expression through NF-[kappa]B activation in vascular endothelial cells. *Atherosclerosis* 185 39-46.

Keavney B 2002 Genetic epidemiological studies of coronary heart disease. *International Journal of Epidemiology* 31 730-736.

Keavney B, Danesh J, Parish S, Palmer A, Clark S, Youngman L, Delepine M, Lathrop M, Peto R, Collins R & The International Studies of Infarct Survival (ISIS) Collaborators 2006 Fibrinogen and coronary heart disease: test of causality by 'Mendelian randomization'. *International Journal of Epidemiology* 35 935-943.

Khera A, McGuire DK, Murphy SA, Stanek HG, Das SR, Vongpatanasin W, Wians J, Grundy SM & de Lemos JA 2005 Race and Gender Differences in C-Reactive Protein Levels. *Journal of the American College of Cardiology* 46 464-469.

Kincer JF, Uittenbogaard A, Dressman J, Guerin TM, Febbraio M, Guo L & Smart EJ 2002 Hypercholesterolemia Promotes a CD36-dependent and Endothelial Nitric-oxide Synthase-mediated Vascular Dysfunction. *Journal of Biological Chemistry* 277 23525-23533.

Kivimaki M, Lawlor DA, Juonala M, Davey Smith G, Elovainio M, Keltikangas-Jarvinen L, Vahtera J, Viikari JSA & Raitakari OT 2005 Lifecourse Socioeconomic Position, C-Reactive Protein, and Carotid Intima-Media Thickness in Young Adults: The Cardiovascular Risk in Young Finns Study. *Arteriosclerosis, Thrombosis, and Vascular Biology* 25 2197-2202.

Klerk M, Verhoef P, Clarke R, Blom HJ, Kok FJ, Schouten EG & and the MTHFR Studies Collaboration Group 2002 MTHFR 677C->T Polymorphism and Risk of Coronary Heart Disease: A Meta-analysis. *JAMA: The Journal of the American Medical Association* 288 2023-2031.

Klimov AN & Nagornev VA 2002 Evolution of cholesterol concept of atherogenesis from Anitchkov to our days. *Pediatr Pathol Mol Med* 21 307-320.

Kluft C & de Maat MPM 2003 Genetics of C-Reactive Protein: New Possibilities and Complications. *Arteriosclerosis, Thrombosis, and Vascular Biology* 23 1956-1959.

Knight JC, Keating BJ, Rockett KA & Kwiatkowski DP 2003 In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat Genet* 33 469-475.

Koenig W, Sund M, Frohlich M, Fischer HG, Lowel H, Doring A, Hutchinson WL & Pepys MB 1999 C-Reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men: results from the MONICA (Monitoring Trends and Determinants in Cardiovascular Disease) Augsburg Cohort Study, 1984 to 1992. *Circulation* 99 237-242.

Koenig W 2005 Predicting risk and treatment benefit in atherosclerosis: the role of C-reactive protein. *International Journal of Cardiology* 98 199-206.

Koenig W, Lowel H, Baumert J & Meisinger C 2004 C-Reactive Protein Modulates Risk Prediction Based on the Framingham Score: Implications for Future Risk Assessment: Results From a Large Cohort Study in Southern Germany. *Circulation* 109 1349-1353.

Koepsell TD & Connell FA 1985 Measures of gain in certainty from a diagnostic test. *American Journal of Epidemiology* 121 744-753.

Kovacs A, Green F, Hansson LO, Lundman P, Samnegard A, Boquist S, Ericsson CG, Watkins H, Hamsten A & Tornvall P 2004 A novel common single nucleotide polymorphism in the promoter region

of the C-reactive protein gene associated with the plasma concentration of C-reactive protein. *Atherosclerosis* 178 193-198.

Kruglyak L 1999 Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 22 139-144.

Kruglyak L 2005 Power tools for human genetics. *Nat Genet* 37 1299-1300.

Kuller LH 2004 Ethnic differences in atherosclerosis, cardiovascular disease and lipid metabolism. *Curr.Opin.Lipidol.* 15 109-113.

Kuller LH, Tracy RP, Shaten J & Meilahn EN 1996 Relation of C-reactive protein and coronary heart disease in the MRFIT nested case-control study. *American Journal of Epidemiology* 144 537-547.

Lakoski SG, Cushman M, Palmas W, Blumenthal R, D'Agostino J & Herrington DM 2005 The Relationship Between Blood Pressure and C-Reactive Protein in the Multi-Ethnic Study of Atherosclerosis (MESA). *Journal of the American College of Cardiology* 46 1869-1874.

Lalla E, Lamster IB, Hofmann MA, Bucciarelli L, Jerud AP, Tucker S, Lu Y, Papapanou PN & Schmidt AM 2003 Oral infection with a periodontal pathogen accelerates early atherosclerosis in apolipoprotein E-null mice. *Arteriosclerosis, Thrombosis, and Vascular Biology* 23 1405-1411.

Lamont RJ & Jenkinson HF 1998 Life below the gum line: pathogenic mechanisms of Porphyromonas gingivalis. *Microbiol.Mol.Biol Rev.* 62 1244-1263.

LaMonte MJ, Durstine JL, Yanowitz FG, Lim T, DuBose KD, Davis P & Ainsworth BE 2002 Cardiorespiratory Fitness and C-Reactive Protein Among a Tri-Ethnic Sample of Women. *Circulation* 106 403-406.

Lao O, Duijn Kv, Kersbergen P, Knijff Pd & Kayser M 2006 Proportioning Whole-Genome Single-Nucleotide-Polymorphism Diversity for the Identification of Geographic Population Structure and Genetic Ancestry. *The American Journal of Human Genetics* 78 680-690.

LaRosa JC, He J & Vupputuri S 1999 Effect of statins on risk of coronary disease: a meta-analysis of randomized controlled trials. *JAMA: The Journal of the American Medical Association* 282 2340-2346.

Law MR, Wald NJ & Morris JK 2004 The performance of blood pressure and other cardiovascular risk factors as screening tests for ischaemic heart disease and stroke. *J Med Screening* 11 3-7.

Lawlor DA, Davey Smith G, Rumley A, Lowe GDO & Ebrahim S 2005 Associations of fibrinogen and C-reactive protein with prevalent and incident coronary heart disease are attenuated by adjustment for confounding factors BritishWomen's Heart and Health Study. *Thromb Haemost.* 93 955-963.

Lawlor DA, Smith GD, Bruckdorfer KR, Kundu D & Ebrahim S 2004 Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? *The Lancet* 363 1724-1727.

Lear SA, Chen MM, Birmingham CL & Frohlich JJ 2003 The relationship between simple anthropometric indices and c-reactive protein: ethnic and gender differences. *Metabolism* 52 1542-1546.

Leelawattana R, Rattarasarn C, Lim A, Soonthornpun S & Setasuban W 2003 Causes of death, incidence and risk factors of cardiovascular diseases in Thai type 2 diabetic patients: a 5 year follow-up study. *Diabetes Res Clin Pract* 60 183-189.

Leon DA, Lithell HO, Vagero D, Koupilova I, Mohsen R, Berglund L, Lithell UB & McKeigue PM 1998 Reduced fetal growth rate and increased risk of death from ischaemic heart disease: cohort study of 15á000 Swedish men and women born 1915-29. *BMJ* 317 241-245.

Leong KH, Koh ET, Feng PH & Boey ML 1994 Lipid profiles in patients with systemic lupus erythematosus. *J Rheumatol.* 21 1264-1267.

Lerner DJ & Kannel WB 1986 Patterns of coronary heart disease morbidity and mortality in the sexes: a 26-year follow-up of the Framingham population. *Am Heart J* 111 383-390.

Lewington S, Clarke R, Qizilbash N, Peto R & Collins R 2002 Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *The Lancet* 360 1903-1913.

Lewis SJ, Ebrahim S & Davey Smith G 2005 Meta-analysis of MTHFR 677C->T polymorphism and coronary heart disease: does totality of evidence support causal role for homocysteine and preventive potential of folate? *BMJ* 331 1053.

Li L, Messas E, Batista EL, Jr., Levine RA & Amar S 2002 Porphyromonas gingivalis infection accelerates the progression of atherosclerosis in a heterozygous apolipoprotein E-deficient murine model. *Circulation* 105 861-867.

Li R, Bensen JT, Hutchinson RG, Province MA, Hertz-Picciotto I, Sprafka JM & Tyroler HA 2000 Family risk score of coronary heart disease (CHD) as a predictor of CHD: the Atherosclerosis Risk in Communities (ARIC) study and the NHLBI family heart study. *Genet Epidemiol.* 18 236-250.

Li SP & Goldman ND 1996 Regulation of human C-reactive protein gene expression by two synergistic IL-6 responsive elements. *Biochemistry* 35 9060-9068.

Libby P, Aikawa M & Schonbeck U 2000 Cholesterol and atherosclerosis. *Biochim.Biophys.Acta* 1529 299-309.

Libby P 1997 Mechanisms of disruption of the atheromatous plaque. *Atherosclerosis* 134 197.

Libby P, Ridker PM & Maseri A 2002 Inflammation and Atherosclerosis. *Circulation* 105 1135-1143.

Lip GYH & Boos CJ 2005 Ethnic Differences in Arterial Responses, Inflammation, and Metabolic Profiles: Possible Insights into Ethnic Differences in Cardiovascular Disease and Stroke. *Arteriosclerosis, Thrombosis, and Vascular Biology* 25 2240-2242.

Livshits G & Gerber LM 2001 Familial Factors of Blood Pressure and Adiposity Covariation. *Hypertension* 37 928-935.

Lowe GD 2001 The relationship between infection, inflammation, and cardiovascular disease: an overview. *Ann.Periodontol.* 6 1-8.

Lowe GDO, Sweetnam PM, Yarnell JWG, Rumley A, Rumley C, Bainton D & Ben Shlomo Y 2004 C-Reactive Protein, Fibrin D-Dimer, and Risk of Ischemic Heart Disease: The Caerphilly and Speedwell Studies. *Arteriosclerosis, Thrombosis, and Vascular Biology* 24 1957-1962.

Lowe GDO, Yarnell JWG, Rumley A, Bainton D & Sweetnam PM 2001 C-Reactive Protein, Fibrin D-Dimer, and Incident Ischemic Heart Disease in the Speedwell Study : Are Inflammation and Fibrin Turnover Linked in Pathogenesis? *Arteriosclerosis, Thrombosis, and Vascular Biology* 21 603-610.

Luc G, Bard JM, Juhan-Vague I, Ferrieres J, Evans A, Amouyel P, Arveiler D, Fruchart JC & Ducimetiere P 2003 C-Reactive Protein, Interleukin-6, and Fibrinogen as Predictors of Coronary Heart Disease: The PRIME Study. *Arterioscler.Thromb.Vasc.Biol.* 23 1255-1261.

Lynch JW, Kaplan GA, Cohen RD, Tuomilehto J & Salonen JT 1996 Do Cardiovascular Risk Factors Explain the Relation between Socioeconomic Status, Risk of All-Cause Mortality, Cardiovascular Mortality, and Acute Myocardial Infarction? *American Journal of Epidemiology* 144 934-942.

MacGregor AJ, Gallimore JR, Spector TD & Pepys MB 2003 Genetic Effects on Baseline Values of C-Reactive Protein and Serum Amyloid A Protein: A Comparison of Monozygotic and Dizygotic Twins. *Clinical Chemistry* 50 130-134.

MacMahon S 2000 Blood Pressure and the Risk of Cardiovascular Disease. *The New England Journal of Medicine* 342 49-52.

MacMahon S, Sharpe N, Gamble G, Clague A, Ni Mhurchu C, Clark T, Hart H, Scott J, White H & Collaborative Research Group 2000 Randomized, placebo-controlled trial of the angiotensin-converting enzyme inhibitor, ramipril, in patients with coronary or other occlusive arterial disease. *Journal of the American College of Cardiology* 36 438-443.

Magyar MT, Szikszai Z, Balla J, Valikovics A, Kappelmayer J, Imre S, Balla G, Jeney V, Csiba L & Bereczki D 2003 Early-onset carotid atherosclerosis is associated with increased intima-media thickness and elevated serum levels of inflammatory markers. *Stroke* 34 58-63.

Majello B, Arcone R, Toniatti C & Ciliberto G 1990 Constitutive and IL-6-induced nuclear factors that interact with the human C-reactive protein promoter. *EMBO J.* 9 457-465.

Maller J, George S, Purcell S, Fagerness J, Altshuler D, Daly MJ & Seddon JM 2006 Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet* 38 1055-1059.

Mamode N, Cobbe S & Pollock JG 1995 Infarcts after surgery. *BMJ* 310 1215-1216.

Marenberg ME, Risch N, Berkman LF, Floderus B & de Faire U 1994 Genetic Susceptibility to Death from Coronary Heart Disease in a Study of Twins. *The New England Journal of Medicine* 330 1041-1046.

Margaglione M, Cappucci G, Colaizzo D, Vecchione G, Grandone E & Di Minno G 2000 C-reactive protein in offspring is associated with the occurrence of myocardial infarction in first-degree relatives. *Arteriosclerosis, Thrombosis, and Vascular Biology* 20 198-203.

Marmot MG, Bosma H, Hemingway H, Brunner E & Stansfeld S 1997 Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *The Lancet* 350 235-239.

Marnell LL, Mold C, Volzer MA, Burlingame RW & Du Clos TW 1995 C-reactive protein binds to Fc[gamma]RI in transfected COS cells. *Journal of Immunology* 155 2185-2193.

Matthews KA, Kelsey SF, Meilahn EN, Muller LH & Wing RR 1989 Educational attainment and behavioral and biologic risk factors for coronary heart disease in middle-aged women. *American Journal of Epidemiology* 129 1132-1144.

Matthews KA, Sowers MF, Derby CA, Stein E, Miracle-McMahill H, Crawford SL & Pasternak RC 2005 Ethnic differences in cardiovascular risk factor burden among middle-aged women: Study of Women's Health Across the Nation (SWAN). *American Heart Journal* 149 1066-1073.

May M, Lawlor DA, Brindle P, Patel R & Ebrahim S 2006 Cardiovascular disease risk assessment in older women - can we improve on Framingham?: British Women's Heart and Health prospective cohort study. *Heart* 92 1396-1401.

McCarthy JJ, Parker A, Salem R, Moliterno DJ, Wang Q, Plow EF, Rao S, Shen G, Rogers WJ, Newby LK, Cannata R, Glatt K & Topol EJ 2004 Large scale association analysis for identification of genes underlying premature coronary heart disease: cumulative perspective from analysis of 111 candidate genes. *Journal of Medical Genetics* 41 334-341.

McGill HC, McMahan CA, Herderick EE, Malcom GT, Tracy RE & Strong JP 2000 Origin of atherosclerosis in childhood and adolescence. *American Journal of Clinical Nutrition* 72 1307S-1315.

McKeigue PM, Shah B & Marmot MG 1991 Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. *The Lancet* 337 382-386.

McNamara DJ 2000 Dietary cholesterol and atherosclerosis. *Biochim.Biophys.Acta* 1529 310-320.

McVean G, Spencer CCA & Chaix R 2005 Perspectives on Human Genetic Variation from the HapMap Project. *PLoS Genetics* 1 e54.

Meade T, Zuhrie R, Cook C & Cooper J 2002 Bezafibrate in men with lower extremity arterial disease: randomised controlled trial. *BMJ* 325 1139.

Meier-Ewert HK, Ridker PM, Rifai N, Price N, Dinges DF & Mullington JM 2001 Absence of diurnal variation of C-reactive protein concentrations in healthy human subjects. *Clinical Chemistry* 47 426-430.

Melnick JL, Adam E & DeBakey ME 1995 Cytomegalovirus and atherosclerosis. *BioEssays: News And Reviews In Molecular, Cellular And Developmental Biology* 17 899-903.

Mendall MA, Strachan DP, Butland BK, Ballam L, Morris J, Sweetnam PM & Elwood PC 2000 C-reactive protein: Relation to total mortality, cardiovascular mortality and cardiovascular risk factors in men. *European Heart Journal* 21 1584-1590.

Mendel G 1865 *Experiments in Plant Hybridization.*

Meredith IT, Currie KE, Anderson TJ, Roddy MA, Ganz P & Creager MA 1996 Postischemic vasodilation in human forearm is dependent on endothelium-derived nitric oxide. *AJP - Heart and Circulatory Physiology* 270 H1435-H1440.

Mezaki T, Matsubara T, Hori T, Higuchi K, Nakamura A, Nakagawa I, Imai S, Ozaki K, Tsuchida K, Nasuno A, Tanaka T, Kubota K, Nakano M, Miida T & Aizawa Y 2003 Plasma levels of soluble thrombomodulin, C-reactive protein, and serum amyloid A protein in the atherosclerotic coronary circulation. *Jpn.Heart J* 44 601-612.

Miller DT, Zee RYL, Suk Danik J, Kozlowski P, Chasman DI, Lazarus R, Cook NR, Ridker PM & Kwiatkowski DJ 2005 Association of Common CRP Gene Variants with CRP Levels and Cardiovascular Events. *Annals of Human Genetics* 69 623-638.

Miller GJ, Bauer KA, Barzegar S, Foley AJ, Mitchell JP, Cooper JA & Rosenberg RD 1995 The effects of quality and timing of venepuncture on markers of blood coagulation in healthy middle-aged men. *Thromb.Haemost.* 73 82-86.

Miller M, Sagnella G, Keser N, Kerry S, Strazzullo P, Cook D & Cappuccio FP 2003 Ethnic differences in circulating soluble adhesion molecules: the Wandsworth Heart and Stroke Study. *Clinical Science* 104 591-598.

Moazed TC, Kuo C, Patton DL, Grayston JT & Campbell LA 1996 Experimental rabbit models of Chlamydia pneumoniae infection. *Am J Pathol.* 148 667-676.

Mohamed-Ali V, Goodrick S, Rawesh A, Katz DR, Miles JM, Yudkin JS, Klein S & Coppack SW 1997 Subcutaneous Adipose Tissue Releases Interleukin-6, But Not Tumor Necrosis Factor-{alpha}, in Vivo. *Journal of Clinical Endocrinology Metabolism* 82 4196-4200.

Morgan TH 1913 *Heredity and sex*. New York: Columbia University Press.

Morrow DA, Rifai N, Antman EM, Weiner DL, McCabe CH, Cannon CP & Braunwald E 1998 C-Reactive Protein Is a Potent Predictor of Mortality Independently of and in Combination With Troponin T in Acute Coronary Syndromes: A TIMI 11A Substudy. *Journal of the American College of Cardiology* 31 1460-1465.

Muhlestein JB, Anderson JL, Hammond EH, Zhao L, Trehan S, Schwobe EP & Carlquist JF 1998 Infection with Chlamydia pneumoniae accelerates the development of atherosclerosis and treatment with azithromycin prevents it in a rabbit model. *Circulation* 97 633-636.

Muir KW, Weir CJ, Alwan W, Squire IB & Lees KR 1999 C-Reactive Protein and Outcome After Ischemic Stroke. *Stroke* 30 981-985.

Munafo MR & Flint J 2004 Meta-analysis of genetic association studies. *Trends in Genetics* 20 439-444.

Murphy C, Beckers J & Ruther U 1995 Regulation of the human C-reactive protein gene in transgenic mice. *Journal of Biological Chemistry* 270 704-708.

Mwantembe O, Gaillard MC, Barkhuizen M, Pillay V, Berry SD, Dewar JB & Song E 2001 Ethnic differences in allelic associations of the interleukin-1 gene cluster in South African patients with inflammatory bowel disease (IBD) and in control individuals. *Immunogenetics* 52 249-254.

Napoli C, de Nigris F & Palinski W 2001 Multiple role of reactive oxygen species in the arterial wall. *J Cell Biochem.* 82 674-682.

Need AC & Goldstein DB 2006 Genome-wide tagging for everyone. *Nat Genet* 38 1227-1228.

Nerurkar SS, McDevitt PJ, Scott GF, Johanson KO, Willette RN & Yue TL 2005 Lipopolysaccharide (LPS) Contamination Plays the Real Role in C-Reactive Protein-Induced IL-6 Secretion From Human Endothelial Cells In Vitro. *Arteriosclerosis, Thrombosis, and Vascular Biology* 25 e136.

Ng SB, Tan YH & Guy GR 1994 Differential induction of the interleukin-6 gene by tumor necrosis factor and interleukin-1. *Journal of Biological Chemistry* 269 19021-19027.

Nora JJ, Lortscher RH, Spangler RD, Nora AH & Kimberling WJ 1980 Genetic--epidemiologic study of early-onset ischemic heart disease. *Circulation* 61 503-508.

O'Connell DL, Heller RF, Roberts DC, Allen JR, Knapp JC, Steele PL & Silove D 1988 Twin study of genetic and environmental effects on lipid levels. *Genet Epidemiol.* 5 323-341.

O'Connor CM, Dunne MW, Pfeffer MA, Muhlestein JB, Yao L, Gupta S, Benner RJ, Fisher MR & Cook TD 2003 Azithromycin for the secondary prevention of coronary heart disease events: the WIZARD

study: a randomized controlled trial. *JAMA: The Journal of the American Medical Association* 290 1459-1466.

Owen N, Poulton T, Hay FC, Mohamed-Ali V & Steptoe A 2003 Socioeconomic status, C-reactive protein, immune factors, and responses to acute mental stress. *Brain, Behavior, and Immunity* 17 286-295.

Packard CJ, O'Reilly DS, Caslake MJ, McMahon AD, Ford I, Cooney J, Macphee CH, Suckling KE, Krishna M, Wilkinson FE, Rumley A & Lowe GD 2000 Lipoprotein-associated phospholipase A2 as an independent predictor of coronary heart disease. West of Scotland Coronary Prevention Study Group. *N.Engl.J.Med.* 343 1148-1155.

Pai JK, Pischon T, Ma J, Manson JE, Hankinson SE, Joshipura K, Curhan GC, Rifai N, Cannuscio CC, Stampfer MJ & Rimm EB 2004 Inflammatory Markers and the Risk of Coronary Heart Disease in Men and Women. *The New England Journal of Medicine* 351 2599-2610.

Panagiotakos DB, Pitsavos CE, Chrysohoou CA, Skoumas J, Toutouza M, Belegrinos D, Toutouzas PK & Stefanadis C 2004 The association between educational status and risk factors related to cardiovascular disease in healthy individuals: The ATTICA study. *Annals of Epidemiology* 14 188-194.

Pankow JS, Folsom AR, Cushman M, Borecki IB, Hopkins PN, Eckfeldt JH & Tracy RP 2001 Familial and genetic determinants of systemic markers of inflammation: the NHLBI family heart study. *Atherosclerosis* 154 681-689.

Paradowski M, Lobos M, Kuydowicz J, Krakowiak M & Kubasiewicz-Ujma B 1995 Acute phase proteins in serum and cerebrospinal fluid in the course of bacterial meningitis. *Clinical Biochemistry* 28 459-466.

Parthasarathy S, Steinberg D & Witztum JL 1992 The role of oxidized low-density lipoproteins in the pathogenesis of atherosclerosis. *Annu.Rev.Med* 43 219-225.

Pasceri V, Willerson JT & Yeh ET 2000 Direct proinflammatory effect of C-reactive protein on human endothelial cells. *Circulation* 102 2165-2168.

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BTN, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SPA & Cox DR 2001 Blocks of Limited Haplotype Diversity Revealed by High-Resolution Scanning of Human Chromosome 21. *Science* 294 1719-1723.

Pearson TA, Blair SN, Daniels SR, Eckel RH, Fair JM, Fortmann SP, Franklin BA, Goldstein LB, Greenland P, Grundy SM, Hong Y, Houston Miller N, Lauer RM, Ockene IS, Sacco RL, Sallis JF, Jr., Smith SC, Jr., Stone NJ & Taubert KA 2002 AHA Guidelines for Primary Prevention of Cardiovascular Disease and Stroke: 2002 Update: Consensus Panel Guide to Comprehensive Risk Reduction for Adult Patients Without Coronary or Other Atherosclerotic Vascular Diseases. *Circulation* 106 388-391.

Pearson TA, Mensah GA, Alexander RW, Anderson JL, Cannon RO, III, Criqui M, Fadl YY, Fortmann SP, Hong Y, Myers GL, Rifai N, Smith SC, Jr., Taubert K, Tracy RP & Vinicor F 2003b Markers of Inflammation and Cardiovascular Disease: Application to Clinical and Public Health Practice: A Statement for Healthcare Professionals From the Centers for Disease Control and Prevention and the American Heart Association. *Circulation* 107 499-511.

Pearson TA, Mensah GA, Alexander RW, Anderson JL, Cannon RO, III, Criqui M, Fadl YY, Fortmann SP, Hong Y, Myers GL, Rifai N, Smith SC, Jr., Taubert K, Tracy RP & Vinicor F 2003a Markers of Inflammation and Cardiovascular Disease: Application to Clinical and Public Health Practice: A Statement for Healthcare Professionals From the Centers for Disease Control and Prevention and the American Heart Association. *Circulation* 107 499-511.

Pedersen TR, Olsson AG, Fµrgeman O, Kjekshus J, Wedel H, Berg K, Wilhelmsen L, Haghfelt T, Thorgeirsson G, Pyorala K, Miettinen T, Christophersen B, Tobert JA, Musliner TA & Cook TJ 1998 Lipoprotein Changes and Reduction in the Incidence of Major Coronary Heart Disease Events in the Scandinavian Simvastatin Survival Study (4S). *Circulation* 97 1453-1460.

Pencina MJ & D'Agostino RB 2004 Overall *C* as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in Medicine* 23 2109-2123.

Pepe MS, Janes H, Longton G, Leisenring W & Newcomb P 2004 Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic, Prognostic, or Screening Marker. *American Journal of Epidemiology* 159 882-890.

Pepys MB 1982 C-reactive protein and the acute phase response. *Nature* 296 12.

Pepys MB, Baltz M, Gomer K, Davies AJS & Doenhoff M 1979 Serum amyloid P-component is an acute-phase reactant in the mouse. *Nature* 278 259-261.

Pepys MB & Hirschfield GM 2001 C-reactive protein and its role in the pathogenesis of myocardial infarction. *Ital.Heart J.* 2 804-806.

Pepys MB & Hirschfield GM 2003 C-reactive protein: a critical update. *J.Clin.Invest* 111 1805-1812.

Pepys MB, Hawkins PN, Kahan MC, Tennent GA, Gallimore JR, Graham D, Sabin CA, Zychlinsky A & de Diego J 2005 Proinflammatory Effects of Bacterial Recombinant Human C-Reactive Protein Are Caused by Contamination With Bacterial Products, Not by C-Reactive Protein Itself. *Circulation Research* 97 e97-103.

Pepys MB, Hirschfield GM, Tennent GA, Ruth Gallimore J, Kahan MC, Bellotti V, Hawkins PN, Myers RM, Smith MD, Polara A, Cobb AJA, Ley SV, Andrew Aquilina J, Robinson CV, Sharif I, Gray GA, Sabin CA, Jenvey MC, Kolstoe SE, Thompson D & Wood SP 2006 Targeting C-reactive protein for the treatment of cardiovascular disease. *Nature* 440 1217-1221.

Phimister EG 2005 Genomic Cartography - Presenting the HapMap. *The New England Journal of Medicine* 353 1766-1768.

Pickup JC & Crook MA 1998 Is type II diabetes mellitus a disease of the innate immune system? *Diabetologia* 41 1241-1248.

Poli V, Mancini FP & Cortese R 1990 IL-6DBP, a nuclear protein involved in interleukin-6 signal transduction, defines a new family of leucine zipper proteins related to C/EBP. *Cell* 63 643-653.

Povoa P, Coelho L, Almeida E, Fernandes A, Mealha R, Moreira P & Sabino H 2005 C-reactive protein as a marker of infection in critically ill patients. *Clinical Microbiology and Infection* 11 101-108.

Pradhan AD, Manson JE, Rossouw JE, Siscovick DS, Mouton CP, Rifai N, Wallace RB, Jackson RD, Pettinger MB & Ridker PM 2002 Inflammatory biomarkers, hormone replacement therapy, and incident coronary heart disease: prospective analysis from the Women's Health Initiative observational study. *JAMA: The Journal of the American Medical Association* 288 980-987.

Qi M, Miyakawa H & Kuramitsu HK 2003 Porphyromonas gingivalis induces murine macrophage foam cell formation. *Microb.Pathog.* 35 259-267.

Qiao JH, Tripathi J, Mishra NK, Cai Y, Tripathi S, Wang XP, Imes S, Fishbein MC, Clinton SK, Libby P, Lusis AJ & Rajavashisth TB 1997 Role of macrophage colony-stimulating factor in atherosclerosis: studies of osteopetrotic mice. *Am.J.Pathol.* 150 1687-1699.

Quandt K, Frech K, Karas H, Wingender E & Werner T 1995 MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* 23 4878-4884.

Raji A, Gerhard-Herman MD, Warren M, Silverman SG, Raptopoulos V, Mantzoros CS & Simonson DC 2004 Insulin Resistance and Vascular Dysfunction in Nondiabetic Asian Indians. *Journal of Clinical Endocrinology Metabolism* 89 3965-3972.

Ramji DP, Vitelli A, Tronche F, Cortese R & Ciliberto G 1993 The two C/EBP isoforms, IL-6DBP/NF-IL6 and C/EBP delta/NF-IL6 beta, are induced by IL-6 to promote acute phase gene transcription via different mechanisms. *Nucleic Acids Res* 21 289-294.

Ray R & Shah AM 2005 NADPH oxidase and endothelial cell function. *Clin Sci.(Lond)* 109 217-226.

Reaven GM 1988 Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes* 37 1595-1607.

Resnick HE, Jones K, Ruotolo G, Jain AK, Henderson J, Lu W & Howard BV 2003 Insulin Resistance, the Metabolic Syndrome, and Risk of Incident Cardiovascular Disease in Nondiabetic American Indians: The Strong Heart Study. *Diabetes Care* 26 861-867.

Retterstol L, Eikvar L & Berg K 2003 A twin study of C-Reactive Protein compared to other risk factors for coronary heart disease. *Atherosclerosis* 169 279-282.

Ricardo AP, Jay S.Kaufman, Kathryn MR, Ana VD-R, Donglin Z & Gerardo H 2007 Early-life and adult socioeconomic status and inflammatory risk markers in adulthood. *European Journal of Epidemiology* 22 55-66.

Rich-Edwards JW, Stampfer MJ, Manson JE, Rosner B, Hankinson SE, Colditz GA, Hennekens CH & Willet WC 1997 Birth weight and risk of cardiovascular disease in a cohort of women followed up since 1976. *BMJ* 315 396-400.

Ridker PM 2003a Clinical application of C-reactive protein for cardiovascular disease detection and prevention. *Circulation* 107 363-369.

Ridker PM 2003b High-sensitivity C-reactive protein and cardiovascular risk: rationale for screening and primary prevention. *Am J Cardiol.* 92 17K-22K.

Ridker PM, Bassuk SS & Toth PP 2003 C-reactive protein and risk of cardiovascular disease: evidence and clinical application. *Curr.Atheroscler.Rep.* 5 341-349.

Ridker PM, Cushman M, Stampfer MJ, Tracy RP & Hennekens CH 1997 Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *N.Engl.J.Med.* 336 973-979.

Ridker PM, Cushman M, Stampfer MJ, Tracy RP & Hennekens CH 1998a Plasma concentration of C-reactive protein and risk of developing peripheral vascular disease. *Circulation* 97 425-428.

Ridker PM & Haughie P 1998 Prospective studies of C-reactive protein as a risk factor for cardiovascular disease. *J.Investig.Med.* 46 391-395.

Ridker PM, Hennekens CH, Buring JE & Rifai N 2000a C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *N.Engl.J.Med.* 342 836-843.

Ridker PM, Rifai N, Rose L, Buring JE & Cook NR 2002 Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. *N.Engl.J.Med.* 347 1557-1565.

Ridker PM, Rifai N, Stampfer MJ & Hennekens CH 2000b Plasma concentration of interleukin-6 and the risk of future myocardial infarction among apparently healthy men. *Circulation* 101 1767-1772.

Ridker PM 2001a High-Sensitivity C-Reactive Protein : Potential Adjunct for Global Risk Assessment in the Primary Prevention of Cardiovascular Disease. *Circulation* 103 1813-1818.

Ridker PM 2001b Role of inflammatory biomarkers in prediction of coronary heart disease. *The Lancet* 358 946-948.

Ridker PM 2003c High-sensitivity C-reactive protein and cardiovascular risk: rationale for screening and primary prevention. *The American Journal of Cardiology* 92 17-22.

Ridker PM, Buring JE, Shih J, Matias M & Hennekens CH 1998b Prospective Study of C-Reactive Protein and the Risk of Future Cardiovascular Events Among Apparently Healthy Women. *Circulation* 98 731-733.

Ridker PM, Hennekens CH, Rifai N, Buring JE & Manson JE 1999 Hormone Replacement Therapy and Increased Plasma Concentration of C-Reactive Protein. *Circulation* 100 713-716.

Ridker PM, Rifai N, Pfeffer M, Sacks F, Lepage S & Braunwald E 2000c Elevation of Tumor Necrosis Factor-{alpha} and Increased Risk of Recurrent Coronary Events After Myocardial Infarction. *Circulation* 101 2149-2153.

Rissanen AM 1979a Familial aggregation of coronary heart disease in a high incidence area (North Karelia, Finland). *Br.Heart J* 42 294-303.

Rissanen AM 1979b Familial occurrence of coronary heart disease: effect of age at diagnosis. *Am.J Cardiol.* 44 60-66.

Rissanen AM 1985 Familial occurrence of coronary heart disease according to clinical manifestation. *Acta Med Scand.* 218 355-363.

Rissanen AM & Nikkila EA 1979 Aggregation of coronary risk factors in families of men with fatal and non-fatal coronary heart disease. *Br.Heart J* 42 373-380.

Rissanen AM & Nikkila EA 1984 Identification of the high-risk groups in familial coronary heart disease. *Atherosclerosis* 53 37-46.

Robert FR 1999 Reference distributions for the negative acute-phase serum proteins, albumin, transferrin and transthyretin: A practical, simple and clinically relevant approach in a large cohort. *Journal of Clinical Laboratory Analysis* 13 273-279.

Roivainen M, Viik-Kajander M, Palosuo T, Toivanen P, Leinonen M, Saikku P, Tenkanen L, Manninen V, Hovi T & Manttari M 2000 Infections, Inflammation, and the Risk of Coronary Heart Disease. *Circulation* 101 252-257.

Roncaglioni MC 2001 Low-dose aspirin and vitamin E in people at cardiovascular risk: a randomised trial in general practice. *The Lancet* 357 89-95.

Rose G 1985 Sick individuals and sick populations. *International Journal of Epidemiology* 14 32-38.

Rosenson RS 2004 Statins in atherosclerosis: lipid-lowering agents with antioxidant capabilities. *Atherosclerosis* 173 1-12.

Ross R 1990 Mechanisms of atherosclerosis--a review. *Adv.Nephrol.Necker Hosp.* 19 79-86.

Ross R 1993 The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature* 362 801-809.

Ross R 1997 Cellular and molecular studies of atherogenesis. *Atherosclerosis* 131 Suppl S3-S4.

Ross R & Agius L 1992 The process of atherogenesis--cellular and molecular interaction: from experimental animal models to humans. *Diabetologia* 35 Suppl 2 S34-S40.

Ross R, Glomset J & Harker L 1977 Response to injury and atherogenesis. *Am J Pathol.* 86 675-684.

Ross R 1999 Atherosclerosis -- An Inflammatory Disease. *The New England Journal of Medicine* 340 115-126.

Ross SD, Allen IE, Connelly JE, Korenblat BM, Smith ME, Bishop D & Luo D 1999 Clinical outcomes in statin treatment trials: a meta-analysis. *Archives of Internal Medicine* 159 1793-1802.

Roumen-Klappe EM, den Heijer M, van Uum SH, van d, V, van der GF & Wollersheim H 2002 Inflammatory response in the acute phase of deep vein thrombosis. *J Vasc Surg.* 35 701-706.

Rubinsztein DC & Easton DF 1999 Apolipoprotein E Genetic Variation and Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders* 10 199-209.

Rudd JHF, Warburton EA, Fryer TD, Jones HA, Clark JC, Antoun N, Johnstrom P, Davenport AP, Kirkpatrick PJ, Arch BN, Pickard JD & Weissberg PL 2002 Imaging Atherosclerotic Plaque Inflammation With [18F]-Fluorodeoxyglucose Positron Emission Tomography. *Circulation* 105 2708-2711.

Rudd JHF, Davies JR & Weissberg PL 2005 Imaging of Atherosclerosis -- Can We Predict Plaque Rupture? *Trends in Cardiovascular Medicine* 15 17-24.

Ruderman NB & Haudenschild C 1984 Diabetes as an atherogenic factor. *Prog.Cardiovasc.Dis.* 26 373-412.

Rus HG, Vlaicu R & Niculescu F 1996 Interleukin-6 and interleukin-8 protein and gene expression in human arterial atherosclerotic wall. *Atherosclerosis* 127 263-271.

Russell AI, Cunninghame Graham DS, Shepherd C, Roberton CA, Whittaker J, Meeks J, Powell RJ, Isenberg DA, Walport MJ & Vyse TJ 2003 Polymorphism at the C-reactive protein locus influences gene expression and predisposes to systemic lupus erythematosus. *Human Molecular Genetics* 13 137-147.

Russell DW, Esser V & Hobbs HH 1989 Molecular basis of familial hypercholesterolemia. *Arteriosclerosis* 9 18-13.

Russell DW, Lehrman MA, Sudhof TC, Yamamoto T, Davis CG, Hobbs HH, Brown MS & Goldstein JL 1986 The LDL receptor in familial hypercholesterolemia: use of human mutations to dissect a membrane protein. *Cold Spring Harb.Symp.Quant.Biol* 51 Pt 2 811-819.

Rutter MK, Meigs JB, Sullivan LM, D'Agostino RB, Sr. & Wilson PWF 2004 C-Reactive Protein, the Metabolic Syndrome, and Prediction of Cardiovascular Events in the Framingham Offspring Study. *Circulation* 110 380-385.

Sadovnick AD, Yee IML & Ebers GC 2000 Factors influencing sib risks for Multiple Sclerosis. *Clinical Genetics* 58 431-435.

Sakkinen P, Abbott RD, Curb JD, Rodriguez BL, Yano K & Tracy RP 2002 C-reactive protein and myocardial infarction. *Journal of Clinical Epidemiology* 55 445-451.

Samani NJ, Thompson JR, O'Toole L, Channer K & Woods KL 1996 A Meta-analysis of the Association of the Deletion Allele of the Angiotensin-Converting Enzyme Gene With Myocardial Infarction. *Circulation* 94 708-712.

Samsioe G 2003 HRT and Cardiovascular Disease. *Annals of the New York Academy of Sciences* 997 358-372.

Sanceau J, Kaisho T, Hirano T & Wietzerbin J 1995 Triggering of the Human Interleukin-6 Gene by Interferon-γ and Tumor Necrosis Factor-alpha in Monocytic Cells Involves Cooperation between Interferon Regulatory Factor-1, NFkappaB, and Sp1 Transcription Factors. *Journal of Biological Chemistry* 270 27920-27931.

Sattar N, McCarey DW, Capell H & McInnes IB 2003 Explaining how "high-grade" systemic inflammation accelerates vascular risk in rheumatoid arthritis. *Circulation* 108 2957-2963.

Sattar N, McConnachie A, O'Reilly D, Upton MN, Greer IA, Smith GD & Watt G 2004 Inverse Association Between Birth Weight and C-Reactive Protein Concentrations in the MIDSPAN Family Study. *Arteriosclerosis, Thrombosis, and Vascular Biology* 24 583-587.

Schmidt AM, Hori O, Jing Xian C, Jian Feng L, Crandall J, Zhang J, Cao R, Shi Du Y, Brett J & Stern D 1995 Advanced glycation endproducts interacting with their endothelial receptor induce expression of vascular cell adhesion molecule-1 (VCAM-1) in cultured human endothelial cells and in mice: A potential mechanism for the accelerated vasculopathy of diabetes. *Journal of Clinical Investigation* 96 1395-1403.

Schulzer M 1994 Diagnostic tests: A statistical review. *Muscle & Nerve* 17 815-819.

Schwartz CJ, Kelley JL, Nerem RM, Sprague EA, Rozek MM, Valente AJ, Edwards EH, Prasad AR, Kerbacher JJ & Logan SA 1989 Pathophysiology of the atherogenic process. *Am J Cardiol.* 64 23G-30G.

Scirica BM, Morrow DA, Verma S, Devaraj S, Jialal I, Scirica BM, Morrow DA, Verma S, Devaraj S & Jialal I 2006 The Verdict Is Still Out. *Circulation* 113 2128-2151.

Seftel HC, Baker SG, Jenkins T & Mendelsohn D 1989 Prevalence of familial hypercholesterolemia in Johannesburg Jews. *Am J Med Genet* 34 545-547.

Seftel HC, Wyndham CH, dos Santos LA, Pilcher GJ & Baker SG 1987 Prevalence of hypercholesterolaemia in young Afrikaners undergoing coronary artery bypass surgery. Ischaemic heart disease risk factors. *S.Afr.Med J* 71 142-144.

Seino Y, Ikeda U, Ikeda M, Yamamoto K, Misawa Y, Hasegawa T, Kano S & Shimada K 1994 Interleukin 6 gene transcripts are expressed in human atherosclerotic lesions. *Cytokine* 6 87-91.

Sesso HD, Buring JE, Rifai N, Blake GJ, Gaziano JM & Ridker PM 2003 C-Reactive Protein and the Risk of Developing Hypertension. *JAMA: The Journal of the American Medical Association* 290 2945-2951.

Shadish W & Haddock C 1994 *The handbook of Research Synthesis.* New York: Russell Sage Foundation.

Shea S, Ottman R, Gabrieli C, Stein Z & Nichols A 1984 Family history as an independent risk factor for coronary artery disease. *J Am.Coll.Cardiol.* 4 793-801.

Shepherd J, Blauw GJ, Murphy MB, Bollen EL, Buckley BM, Cobbe SM, Ford I, Gaw A, Hyland M, Jukema JW, Kamper AM, Macfarlane PW, Meinders AE, Norrie J, Packard CJ, Perry IJ, Stott DJ, Sweeney BJ, Twomey C & Westendorp RG 2002 Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. *Lancet* 360 1623-1630.

Shlipak MG, Fried LF, Cushman M, Manolio TA, Peterson D, Stehman-Breen C, Bleyer A, Newman A, Siscovick D & Psaty B 2005 Cardiovascular Mortality Risk in Chronic Kidney Disease: Comparison of Traditional and Novel Risk Factors. *JAMA: The Journal of the American Medical Association* 293 1737-1745.

Shohet RV, Vega GL, Bersot TP, Mahley RW, Grundy SM, Guerra R & Cohen JC 2002 Sources of variability in genetic association studies: Insights from the analysis of hepatic lipase (LIPC). *Human Mutation* 19 536-542.

Sing CF & Davignon J 1985 Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. *Am.J Hum.Genet* 37 268-285.

Sinha S, Rathi M, Misra A, Kumar V, Kumar M, Jagannathan NR, Pandey RM, Dwivedi M & Luthra K 2005 Subclinical inflammation and soleus muscle intramyocellular lipids in healthy Asian Indian males. *Clinical Endocrinology* 63 350-355.

Sitzer M, Markus HS, Mendall MA, Liehr R, Knorr U & Steinmetz H 2002 C-reactive protein and carotid intimal medial thickness in a community population. *J.Cardiovasc.Risk* 9 97-103.

Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P & Vallance P 2004 Risk of Myocardial Infarction and Stroke after Acute Infection or Vaccination. *The New England Journal of Medicine* 351 2611-2618.

Snowden CB, McNamara PM, Garrison RJ, Feinleib M, Kannel WB & Epstein FH 1982 Predicting coronary heart disease in siblings--a multivariate assessment: the Framingham Heart Study. *Am.J Epidemiol.* 115 217-222.

Song Y, Stampfer MJ & Liu S 2004 Meta-Analysis: Apolipoprotein E Genotypes and Risk for Coronary Heart Disease. *Annals of Internal Medicine* 141 137-147.

Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ & Cheung VG 2007 Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* 39 226-231.

St Pierre AC, Cantin B, Bergeron J, Pirro M, Dagenais GR, Jean P & Lamarche B 2005 Inflammatory markers and long-term risk of ischemic heart disease in men: A 13-year follow-up of the Quebec Cardiovascular Study. *Atherosclerosis* 182 315-321.

Staessen JA, Wang JG, Ginocchio G, Petrov V, Saavedra AP, Soubrier F, Vlietinck R & Fagard R 1997 The deletion/insertion polymorphism of the angiotensin converting enzyme gene and cardiovascular-renal risk. *J Hypertens.* 15 1579-1592.

Stamler J, Neaton JD & Wentworth DN 1989 Blood pressure (systolic and diastolic) and risk of fatal coronary heart disease. *Hypertension* 13 I2-12.

Stamler J, Stamler R & Neaton JD 1993 Blood pressure, systolic and diastolic, and cardiovascular risks. US population data. *Archives of Internal Medicine* 153 598-615.

Stary HC 1989 Evolution and progression of atherosclerotic lesions in coronary arteries of children and young adults. *Arteriosclerosis* 9 I19-I32.

Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han JH, Duan J, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windemuth A, Xu C, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G & Vovis GF 2001a Haplotype Variation and Linkage Disequilibrium in 313 Human Genes. *Science* 293 489-493.

Stephens M, Smith NJ & Donnelly P 2001b A new statistical method for haplotype reconstruction from population data. *Am J Hum.Genet* 68 978-989.

Stoltenberg JL, Osborn JB, Pihlstrom BL, Herzberg MC, Aeppli DM, Wolff LF & Fischer GE 1993 Association between cigarette smoking, bacterial pathogens, and periodontal status. *J Periodontol.* 64 1225-1230.

Stone PH, Coskun AU, Yeghiazarians Y, Kinlay S, Popma JJ, Kuntz RE & Feldman CL 2003 Prediction of sites of coronary atherosclerosis progression: In vivo profiling of endothelial shear stress, lumen, and outer vessel wall characteristics to predict vascular behavior. *Curr.Opin.Cardiol.* 18 458-470.

Stork S, Feelders RA, van den Beld AW, Steyerberg EW, Savelkoul HFJ, Lamberts SWJ, Grobbee DE & Bots ML 2006 Prediction of Mortality Risk in the Elderly. *The American Journal of Medicine* 119 519-525.

Strandberg TE & Tilvis RS 2000 C-Reactive Protein, Cardiovascular Risk Factors, and Mortality in a Prospective Study in the Elderly. *Arteriosclerosis, Thrombosis, and Vascular Biology* 20 1057-1060.

Suk Danik J, Chasman DI, Cannon CP, Miller DT, Zee RYL, Kozlowski P, Kwiatkowski DJ & Ridker PM 2006 Influence of Genetic Variation in the C-Reactive Protein Gene on the Inflammatory Response During and After Acute Coronary Ischemia. *Annals of Human Genetics* 70 705-716.

Suk HJ, Ridker PM, Cook NR & Zee RYL 2006 Relation of polymorphism within the C-reactive protein gene and plasma CRP levels. *Atherosclerosis* 70 705-716.

Swenson CJ, Trepka MJ, Rewers MJ, Scarbro S, Hiatt WR & Hamman RF 2002 Cardiovascular Disease Mortality in Hispanics and Non-Hispanic Whites. *American Journal of Epidemiology* 156 919-928.

Szalai AJ 2002a The antimicrobial activity of C-reactive protein. *Microbes and Infection* 4 201-205.

Szalai AJ, McCrory MA, Cooper GS, Wu J & Kimberly RP 2002 Association between baseline levels of C-reactive protein (CRP) and a dinucleotide repeat polymorphism in the intron of the CRP gene. *Genes Immun.* 3 14-19.

Szalai AJ, Van Ginkel FW, Wang Y, McGhee JR & Volanakis JE 2000a Complement-dependent acute-phase expression of C-reactive protein and serum amyloid P-component. *Journal of Immunology* 165 1030-1035.

Szalai AJ, Wu J, Lange EM, McCrory MA, Langefeld CD, Williams A, Zakharkin SO, George V, Allison DB, Cooper GS, Xie F, Fan Z, Edberg JC & Kimberly RP 2005 Single-nucleotide polymorphisms in the C-reactive protein (CRP) gene promoter that affect transcription factor binding, alter transcriptional activity, and associate with differences in baseline serum CRP level. *Journal of Molecular Medicine* 83 440-447.

Szalai AJ 2002b The biological functions of C-reactive protein. *Vascular Pharmacology* 39 105-107.

Szalai AJ, VanCott JL, McGhee JR, Volanakis JE & Benjamin WH, Jr. 2000b Human C-Reactive Protein Is Protective against Fatal Salmonella enterica Serovar Typhimurium Infection in Transgenic Mice. *Infection and Immunity* 68 5652-5656.

Tahvanainen E, Syvanne M, Frick MH, Murtomaki-Repo S, Antikainen M, Kesaniemi YA, Kauma H, Pasternak A, Taskinen MR & Ehnholm C 1998 Association of Variation in Hepatic Lipase Activity with Promoter Variation in the Hepatic Lipase Gene. *Journal of Clinical Investigation* 101 956-960.

Talmud PJ, Berglund L, Hawe EM, Waterworth DM, Isasi CR, Deckelbaum RE, Starc T, Ginsberg HN, Humphries SE & Shea S 2001 Age-Related Effects of Genetic Variation on Lipid Levels: The Columbia University BioMarkers Study. *Pediatrics* 108 e50.

Talmud PJ, Hawe E, Martin S, Olivier M, Miller GJ, Rubin EM, Pennacchio LA & Humphries SE 2002 Relative contribution of variation within the APOC3/A4/A5 gene cluster in determining plasma triglycerides. *Human Molecular Genetics* 11 3039-3046.

Tang H, Quertermous T, Rodriguez B, Kardia SL, Zhu X, Brown A, Pankow JS, Province MA, Hunt SC, Boerwinkle E, Schork NJ & Risch NJ 2005 Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *The American Journal of Human Genetics* 76 268-275.

Tang QQ, Otto TC & Lane MD 2003 CCAAT/enhancer-binding protein beta is required for mitotic clonal expansion during adipogenesis. *Proceedings of the National Academy of Sciences* 100 850-855.

Tellides G, Tereb DA, Kirkiles-Smith NC, Kim RW, Wilson JH, Schechner JS, Lorber MI & Pober JS 2000 Interferon-gamma elicits arteriosclerosis in the absence of leukocytes. *Nature* 403 207-211.

The Heart Outcomes Prevention Evaluation Study Investigators 2000 Effects of an Angiotensin-Converting-Enzyme Inhibitor, Ramipril, on Cardiovascular Events in High-Risk Patients. *The New England Journal of Medicine* 342 145-153.

The International HapMap Consortium 2005 A haplotype map of the human genome. *Nature* 437 1299-1320.

Thompson D, Pepys MB & Wood SP 1999 The physiological structure of human C-reactive protein and its complex with phosphocholine. *Structure.Fold.Des* 7 169-177.

Thompson J, Durham L, Lira M, Shear C & Milos P 2005 CETP polymorphisms associated with HDL cholesterol may differ from those associated with cardiovascular disease. *Atherosclerosis* 181 45-53.

Tillet WS & Francis T 1930 Serological reactions in pneumonia with a non-protein somatic fraction of the Pneumococcus. *J Exp.Med* 52 561-571.

Tillin T, Forouhi N, Johnston D, McKeigue P, Chaturvedi N & Godsland I 2005 Metabolic syndrome and coronary heart disease in South Asians, African-Caribbeans and white Europeans: a UK population-based cross-sectional study. *Diabetologia* 48 649-656.

Timpson NJ, Lawlor DA, Harbord RM, Gaunt TR, Day IN, Palmer LJ, Hattersley AT, Ebrahim S, Lowe GD, Rumley A & Smith GD 2005 C-reactive protein and its role in metabolic syndrome: mendelian randomisation study. *The Lancet* 366 1954-1959.

Tobin MD, Minelli C, Burton PR & Thompson JR 2004 Commentary: Development of Mendelian randomization: from hypothesis test to 'Mendelian deconfounding'. *International Journal of Epidemiology* 33 26-29.

Toniatti C, Arcone R, Majello B, Ganter U, Arpaia G & Ciliberto G 1990a Regulation of the human C-reactive protein gene, a major marker of inflammation and cancer. *MOL.BIOL.MED.* 7 199-212.

Toniatti C, Demartis A, Monaci P, Nicosia A & Ciliberto G 1990b Synergistic trans-activation of the human C-reactive protein promoter by transcription factor HNF-1 binding at two distinct sites. *EMBO J.* 9 4467-4475.

Tracy RP 1999 Inflammation markers and coronary heart disease. *Curr.Opin.Lipidol.* 10 435-441.

Tracy RP, Lemaitre RN, Psaty BM, Ives DG, Evans RW, Cushman M, Meilahn EN & Kuller LH 1997 Relationship of C-Reactive Protein to Risk of Cardiovascular Disease in the Elderly: Results From the Cardiovascular Health Study and the Rural Health Promotion Project. *Arteriosclerosis, Thrombosis, and Vascular Biology* 17 1121-1127.

Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas AM & Pajak A 1994 Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. *Circulation* 90 583-612.

Turnbull B 1995 Smoking and periodontal disease. A review. *J N Z.Soc.Periodontol.* 10-15.

Tybjarg-Hansen A, Steffensen R, Meinertz H, Schnohr P & Nordestgaard BG 1998 Association of Mutations in the Apolipoprotein B Gene with Hypercholesterolemia and the Risk of Ischemic Heart Disease. *The New England Journal of Medicine* 338 1577-1584.

Tzoulaki I, Murray GD, Lee AJ, Rumley A, Lowe GDO & Fowkes FG 2005 C-Reactive Protein, Interleukin-6, and Soluble Adhesion Molecules as Predictors of Progressive Peripheral Atherosclerosis in the General Population: Edinburgh Artery Study. *Circulation* 112 976-983.

Ueshima H, Okayama A, Saitoh S, Nakagawa H, Rodriguez B, Sakata K, Okuda N, Choudhury SR & Curb JD 2003 Differences in cardiovascular disease risk factors between Japanese in Japan and Japanese-Americans in Hawaii: the INTERLIPID study. *Journal of Human Hypertension* 17 631-639.

Van Den Berg CW, Taylor KE & Lang D 2004 C-Reactive Protein-Induced In Vitro Vasorelaxation Is an Artefact Caused by the Presence of Sodium Azide in Commercial Preparations. *Arteriosclerosis, Thrombosis, and Vascular Biology.* 24 e168-e171.

van der Meer I, de Maat MP, Kiliaan AJ, van der Kuip DA, Hofman A & Witteman JC 2003 The value of C-reactive protein in cardiovascular risk prediction: the Rotterdam Study. *Arch.Intern.Med.* 163 1323-1328.

Vega GL, Clark LT, Tang A, Marcovina S, Grundy SM & Cohen JC 1998 Hepatic lipase activity is lower in African American men than in white American men: effects of 5' flanking polymorphism in the hepatic lipase gene (LIPC). *Journal of Lipid Research* 39 228-232.

Velican C & Velican D 1985 Differences in the pattern of atherosclerotic involvement between non-branched regions and adjacent branching points of human coronary arteries. *Atherosclerosis* 54 333-342.

Velican D, Petrescu C & Velican C 1986 Atherogenic anatomical patterns of coronary arteries branching related to cardiac and non-cardiac causes of death. *Med Interne* 24 103-109.

Venugopal SK, Devaraj S, Yuhanna I, Shaul P & Jialal I 2002 Demonstration that C-reactive protein decreases eNOS expression and bioactivity in human aortic endothelial cells. *Circulation* 106 1439-1441.

Verkerk MSA, Visseren FLJ, Paul Bouter K & Diepersloot RJA 2003 Acute-phase response of human hepatocytes after infection with Chlamydia pneumoniae and cytomegalovirus. *European Journal of Clinical Investigation* 33 720-725.

Verma S, Szmitko PE & Ridker PM 2005 C-reactive protein comes of age. *Nature Clinical Practice Cardiovascular Medicine* 2 29-36.

Verma S, Wang CH, Li SH, Dumont AS, Fedak PWM, Badiwala MV, Dhillon B, Weisel RD, Li RK, Mickle DAG & Stewart DJ 2002 A Self-Fulfilling Prophecy: C-Reactive Protein Attenuates Nitric Oxide Production and Inhibits Angiogenesis. *Circulation* 106 913-919.

Vita JA, Treasure CB, Nabel EG, McLenachan JM, Fish RD, Yeung AC, Vekshtein VI, Selwyn AP & Ganz P 1990 Coronary vasomotor response to acetylcholine relates to risk factors for coronary artery disease. *Circulation* 81 491-497.

Vliegen I, Duijvestijn A, Grauls G, Herngreen S, Bruggeman C & Stassen F 2004 Cytomegalovirus infection aggravates atherogenesis in apoE knockout mice by both local and systemic immune activation. *Microbes.Infect.* 6 17-24.

Volanakis JE 1982a Complement activation by C-reactive protein complexes. *ANN.N.Y.ACAD.SCI.* Vol. 389 235-250.

Volanakis JE 1982b Complement-induced solubilization of C-reactive protein-pneumococcal C-polysaccharide precipitates: Evidence for covalent binding of complement proteins to C-reactive protein and to pneumococcal C-polysaccharide. *J.Immunol.* 128 2745-2750.

Volanakis JE & Kaplan MH 1971 Specificity of C-reactive protein for choline phosphate residues of pneumococcal C-polysaccharide. *Proceedings Of The Society For Experimental Biology And Medicine.Society For Experimental Biology And Medicine (New York, N.Y.)* 136 612-614.

Volanakis JE 2001 Human C-reactive protein: expression, structure, and function. *Molecular Immunology* 38 189-197.

Vormittag R, Funk M, Mannhalter C, Schonauer V, Vukovich T, Minar E, Bialonczyk C, Hirschl M & Pabinger I 2006 C-reactive protein 3' UTR +1444C>T polymorphism in patients with spontaneous venous thromboembolism. *Atherosclerosis* 188 406-411.

Vos J, Aarnoudse MW, Dijk F & Lamberts HB 1983 On the cellular origin and development of atheromatous plaques. A light and electron microscopic study of combined X-ray and hypercholesterolemia-induced atheromatosis in the carotid artery of the rabbit. *Virchows Arch B Cell Pathol.Incl.Mol.Pathol.* 43 1-16.

Vrecer M, Turk S, Drinovec J & Mrhar A 2003 Use of statins in primary and secondary prevention of coronary heart disease and ischemic stroke. Meta-analysis of randomized trials. *Int.J Clin Pharmacol.Ther.* 41 567-577.

Wajchenberg BL, Malerbi DA, Rocha MS, Lerario AC & Santomauro AT 1994 Syndrome X: a syndrome of insulin resistance. Epidemiological and clinical evidence. *Diabetes Metab Rev.* 10 19-29.

Wald DS, Law M & Morris JK 2002 Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. *BMJ* 325 1202-1206.

Wald DS, Wald NJ, Morris JK & Law M 2006 Folic acid, homocysteine, and cardiovascular disease: judging causality in the face of inconclusive trial evidence. *BMJ* 333 1114-1117.

Wald NJ, Hackshaw AK & Frost CD 1999 When can a risk factor be used as a worthwhile screening test? *BMJ* 319 1562-1565.

Wald NJ & Law MR 2003 A strategy to reduce cardiovascular disease by more than 80%. *BMJ* 326 1419.

Wald NJ, Law MR, Watt HC, Wu T, Bailey A, Johnson AM, Craig WY, Ledue TB & Haddow JE 1994 Apolipoproteins and ischaemic heart disease: implications for screening. *Lancet* 343 75-79.

Wald NJ, Morris JK & Rish S 2005 The efficacy of combining several risk factors as a screening test. *Journal of Medical Screening* 12 197-201.

Wareham NJ, Byrne CD, Williams R, Day NE & Hales CN 1999 Fasting proinsulin concentrations predict the development of type 2 diabetes. *Diabetes Care* 22 262-270.

Warner SJ & Libby P 1989 Human vascular smooth muscle cells. Target for and source of tumor necrosis factor. *J.Immunol.* 142 100-109.

Warren HS & Chedid LA 1987 Strategies for the treatment of endotoxemia: significance of the acute-phase response. *Rev Infect Dis* 9 S630-S638.

Weber JL, Kwitek AE & May PE 1990 Dinucleotide repeat polymorphism at the CRP locus. *Nucleic Acids Res* 18 4635.

Welty TK, Rhoades DA, Yeh F, Lee ET, Cowan LD, Fabsitz RR, Robbins DC, Devereux RB, Henderson JA & Howard BV 2002 Changes in Cardiovascular Disease Risk Factors among American Indians: The Strong Heart Study. *Annals of Epidemiology* 12 97-106.

Wener MH, Daum PR & McQuillan GM 2000 The influence of age, sex, and race on the upper reference limit of serum C-reactive protein concentration. *J Rheumatol.* 27 2351-2359.

Wild SH, Laws A, Fortmann SP, Varady AN & Byrne CD 1995 Mortality from coronary heart disease and stroke for six ethnic groups in California, 1985 to 1990. *Annals of Epidemiology* 5 432-439.

Williams RR, Hunt SC, Hopkins PN, Wu LL, Hasstedt SJ, Berry TD, Barlow GK, Stults BM, Schumacher MC, Ludwig EH & . 1993 Genetic basis of familial dyslipidemia and hypertension: 15-year results from Utah. *Am.J Hypertens.* 6 319S-327S.

Willot S, Vermeire S, Ohresser M, Rutgeerts P, Paintaud G, Belaiche J, De Vos M, Van Gossum A, Franchimont D, Colombel JF, Watier H & Louis E 2006 No association between C-reactive protein gene polymorphisms and decrease of C-reactive protein serum concentration after infliximab treatment in Crohn's disease. *Pharmacogenetics & Genomics* 16 37-42.

Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H & Kannel WB 1998 Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation* 97 1837-1847.

Wilson PWF, Nam BH, Pencina M, D'Agostino RB, Sr., Benjamin EJ & O'Donnell CJ 2005 C-Reactive Protein and Risk of Cardiovascular Disease in Men and Women From the Framingham Heart Study. *Archives of Internal Medicine* 165 2473-2478.

Winbeck K, Kukla C, Poppert H, Klingelh÷fer J, Conrad B & Sander D 2002 Elevated C-Reactive Protein Is Associated with an Increased Intima to Media Thickness of the Common Carotid Artery. *Cerebrovascular.Diseases.* 13 57-63.

Wingard DL, Suarez L & Barrett-Connor E 1983 The sex differential in mortality from all causes and ischemic heart disease. *Am J Epidemiol.* 117 165-172.

Winyard PG, Tatzber F, Esterbauer H, Kus ML, Blake DR & Morris CJ 1993 Presence of foam cells containing oxidised low density lipoprotein in the synovial membrane from patients with rheumatoid arthritis. *Ann.Rheum.Dis.* 52 677-680.

Witztum JL & Steinberg D 1991 Role of oxidized low density lipoprotein in atherogenesis. *J Clin Invest* 88 1785-1792.

Wolford JK, Gruber JD, Ossowski VM, Vozarova B, Antonio Tataranni P, Bogardus C & Hanson RL 2003 A C-reactive protein promoter polymorphism is associated with type 2 diabetes mellitus in Pima Indians. *Molecular Genetics and Metabolism* 78 136-144.

Woloshin S & Schwartz LM 2005 Distribution of C-Reactive Protein Values in the United States. *The New England Journal of Medicine* 352 1611-1613.

Wong TY, Islam FMA, Klein R, Klein BEK, Cotch MF, Castro C, Sharrett AR & Shahar E 2006 Retinal Vascular Caliber, Cardiovascular Risk Factors, and Inflammation: The Multi-Ethnic Study of Atherosclerosis (MESA). *Investigative Ophthalmology Visual Science* 47 2341-2350.

Woo KS & Donnan SPB 1989 Epidemiology of coronary arterial disease in the Chinese. *International Journal of Cardiology* 24 83-93.

Woodhouse PR, Meade TW & Khaw KT 1997 Plasminogen activator inhibitor-1, the acute phase response and vitamin C. *Atherosclerosis* 133 71-76.

Worns MA, Victor A, Galle PR & Hohler T 2006 Genetic and environmental contributions to plasma C-reactive protein and interleukin-6 levels - a study in twins. *Genes Immun* 7 600-605.

WOSCOPS Group 1997 Baseline risk factors and their association with outcome in the West of Scotland Coronary Prevention Study. The West of Scotland Coronary Prevention Study Group. *Am.J.Cardiol.* 79 756-762.

Wyndham CH, Seftel HC, Pilcher GJ & Baker SG 1987 Prevalence of hypercholesterolaemia in young Afrikaners with myocardial infarction. Ischaemic heart disease risk factors. *S.Afr.Med J* 71 139-142.

Xia D & Samols D 1997 Transgenic mice expressing rabbit C-reactive protein are resistant to endotoxemia. *Proceedings of the National Academy of Sciences* 94 2575-2580.

Xu N, Chen CY & Shyu AB 1997 Modulation of the fate of cytoplasmic mRNA by AU-rich elements: key sequence features controlling mRNA deadenylation and decay. *Molecular and Cellular Biology* 17 4611-4621.

Yamada S, Gotoh T, Nakashima Y, Kayaba K, Ishikawa S, Nago N, Nakamura Y, Itoh Y & Kajii E 2001 Distribution of serum C-reactive protein and its association with atherosclerotic risk factors in a Japanese population : Jichi Medical School Cohort Study. *American Journal of Epidemiology* 153 1183-1190.

Ye Z, Liu EH, Higgins JP, Keavney BD, Lowe GD, Collins R & Danesh J 2006 Seven haemostatic gene polymorphisms in coronary disease: meta-analysis of 66,155 cases and 91,307 controls. *The Lancet* 367 651-658.

Yeh ET & Willerson JT 2003 Coming of age of C-reactive protein: using inflammation markers in cardiology. *Circulation* 107 370-371.

Yildirim K, Karatay S, Melikoglu MA, Gureser G, Ugur M & Senel K 2004 Associations between Acute Phase Reactant Levels and Disease Activity Score (DAS28) in Patients with Rheumatoid Arthritis. *Annals of Clinical and Laboratory Science* 34 423-426.

Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, McQueen M, Budaj A, Pais P, Varigos J & Lisheng L 2004 Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *The Lancet* 364 937-952.

Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI & de Faire U 2002 Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *Journal of Internal Medicine* 252 247-254.

Zebrack JS, Anderson JL, Maycock CA, Horne BD, Bair TL & Muhlestein JB 2002 Usefulness of high-sensitivity C-Reactive protein in predicting long-term risk of death or acute myocardial infarction in patients with unstable or stable angina pectoris or acute myocardial infarction. *The American Journal of Cardiology* 89 145-149.

Zee RYL & Ridker PM 2002 Polymorphism in the human C-reactive protein (CRP) gene, plasma concentrations of CRP, and the risk of future arterial thrombosis. *Atherosclerosis* 162 217-219.

Zhang D, Sun M, Samols D & Kushner I 1996 STAT3 participates in transcriptional activation of the C-reactive protein gene by interleukin-6. *Journal of Biological Chemistry* 271 9503-9509.

Zhang K, Shen X, Wu J, Sakaki K, Saunders T, Rutkowski DT, Back SH & Kaufman RJ 2006a Endoplasmic Reticulum Stress Activates Cleavage of CREBH to Induce a Systemic Inflammatory Response. *Cell* 124 587-599.

Zhang Y, Lee ET, Devereux RB, Yeh J, Best LG, Fabsitz RR & Howard BV 2006b Prehypertension, Diabetes, and Cardiovascular Disease Risk in a Population-Based Sample: The Strong Heart Study. *Hypertension* 47 410-414.

Zhou X, Nicoletti A, Elhage R & Hansson GK 2000 Transfer of CD4(+) T cells aggravates atherosclerosis in immunodeficient apolipoprotein E knockout mice. *Circulation* 102 2919-2922.

Zondervan KT & Cardon LR 2004 The complex interplay among factors that influence allelic association. *Nature Reviews Genetics* 5 89-100.

Zoratti R 1998 A review on ethnic differences in plasma triglycerides and high-density-lipoprotein cholesterol: Is the lipid pattern the key factor for the low coronary heart disease rate in people of African origin? *European Journal of Epidemiology* 14 9-21.

Zoratti R, Godsland IF, Chaturvedi N, Crook D, Stevenson JC & McKeigue PM 2000 Relation of plasma lipids to insulin resistance, nonesterified fatty acid levels, and body fat in men from three ethnic groups: Relevance to variation in risk of diabetes and coronary disease. *Metabolism* 49 245-252.